

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 00:16:43 ; Search time 2677 Seconds
(without alignments)
17907.598 Million cell updates/sec

Title: US-09-734-329-1
Perfect score: 2960
Sequence: 1 atttccattctctcc.....aaaaaaaaaaaaaaaaaaaaa 2960

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0
rched: 16154066 seqs, 8097743376 residues

Word size : 0
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:
1: em_estba:*
2: em_esthm:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_esti:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pin:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rdo:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	529	17.9	970	9	AV227235
2	502	17.0	515	9	AA034571
3	490	16.6	615	10	BB622399
4	485	16.4	707	10	BB028252
5	405	13.7	637	10	BB048324
6	388	13.1	643	10	BB622895

7	326	11.0	430	10	BB675901
8	273	9.2	428	10	BB674942
9	269	9.1	670	10	BB618195
10	267	9.0	652	10	AV337525
11	240	8.1	155	10	AW743495
12	152	5.1	155	17	AZ588187
13	111	3.8	296	10	BB527673
14	102	3.4	435	10	BB689394
15	101	3.4	164	17	AZ736987
16	84	2.8	465	12	BF564310
17	84	2.8	520	12	BF564272
18	84	2.8	562	13	BM382845
19	84	2.8	594	13	BM382845
20	84	2.8	661	14	BQ205022
21	84	2.8	681	13	BM389110
22	75	2.5	280	10	BB360632
23	73	2.5	289	10	BB375475
24	69	2.3	282	10	BB521809
25	67	2.3	307	10	BB132968
26	59	2.0	290	10	BB470471
27	59	2.0	322	10	BB363167
28	57	1.9	300	10	BB175776
29	55	1.9	305	10	AW325923
30	54	1.8	260	10	BB523022
31	54	1.8	290	10	BB368597
32	54	1.8	298	10	BB526105
33	54	1.8	301	10	BB133795
34	48	1.6	315	9	AV235790
35	46	1.6	298	10	BB525281
36	43	1.5	339	10	BB131724
37	41	1.4	247	10	BB595005
38	40	1.4	319	10	BB374512
39	39	1.3	252	9	AV245678
40	39	1.3	285	9	AV204031
41	39	1.3	306	14	BQ798667
42	39	1.3	425	10	BE638807
43	38	1.3	255	10	BE552541
44	38	1.3	313	13	BI307226
45	38	1.3	648	14	BM992111

ALIGNMENTS

RESULT 1
AV227235
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

970 bp mRNA linear EST 14-NOV-2001
musculus full-length enriched, 16 days embryo head Mus
musculus cDNA clone 4122402C11 3', mRNA sequence.
GI:16385425

AV227235
RIKEN
EST.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 970)
Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanaoka, T., Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
Unpublished (2001)
On Nov 1, 1999 this sequence version replaced gi:16178902.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222

102(b)

RESULT 2
AA034571
LOCUS
DEFINITION
AA034571 515 bp mRNA linear EST 23-AUG-1996
clone1191.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA
mRNA IMAGE:466845 5', mRNA sequence.
ACCESSION
AA034571
VERSION
AA034571.1 GI:1506434
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 515)
REFERENCE
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
TITLE
The WashU-RHMI Mouse EST Project
JOURNAL
Unpublished (1996)
COMMENT
Contact: Marra M/Mouse EST Project
WashU-RHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:280661
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 505.
Location/Qualifiers
1..515
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:466845"

FEATURES
source

TITLE	JOURNAL	COMMENT
1. The Role of the State in the Development of the Economy	Journal of Economic Surveys	See also 2.
2. The Role of the State in the Development of the Economy	Journal of Economic Surveys	See also 1.
3. The Role of the State in the Development of the Economy	Journal of Economic Surveys	See also 1, 2.
4. The Role of the State in the Development of the Economy	Journal of Economic Surveys	See also 1, 2, 3.
5. The Role of the State in the Development of the Economy	Journal of Economic Surveys	See also 1, 2, 3, 4.
6. The Role of the State in the Development of the Economy	Journal of Economic Surveys	See also 1, 2, 3, 4, 5.
7. The Role of the State in the Development of the Economy	Journal of Economic Surveys	See also 1, 2, 3, 4, 5, 6.
8. The Role of the State in the Development of the Economy	Journal of Economic Surveys	See also 1, 2, 3, 4, 5, 6, 7.
9. The Role of the State in the Development of the Economy	Journal of Economic Surveys	See also 1, 2, 3, 4, 5, 6, 7, 8.
10. The Role of the State in the Development of the Economy	Journal of Economic Surveys	See also 1, 2, 3, 4, 5, 6, 7, 8, 9.

Location/Qualifiers

BASE COUNT


```
Db 130 CCTTGCTCGTAGATTCCTATCTCGAGGCTCCGAGAGTTCTTTTTCAGTTGAGTTTG 189
QY 2241 GGTGTGCGCCCTTTTGTAGAGTTTCTGGGGTGTCTCTGTGTAGCAGTCACTAAGAT 2300
Db 190 GGTGTGCGCCCTTTTAAAGTTTCTGTGGGTGTCTCTGTGTAGCAGTCACTAAGAT 249
QY 2301 CCCAGCCCCAGCAGAAAGCTGTGAAGCTTCAAGTCCCTATGCGGGGAGGACTGGAATG 2360
Db 250 CCCAGCCCCAGCAGAAAGCTGTGAAGCTTCAAGTCCCTATGCGGGGAGGACTGGAATG 309
QY 2361 TACCCAGTCTCTCGACCCGACTGCAGATCAGGTTCTCTCCCTGTATCCTTCTTCATAC 2420
Db 310 TACCCAGTCTCTCGACCCGACTGCAGATCAGGTTCTCTCCCTGTATCCTTCTTCATAC 369
QY 2421 CTTGTGACCTCACAGGTTATCCCTTGTGTCATGCTGTTACAGAGAGCTTGACAGTGCCA 2480
Db 370 CTTGTGACCTCACAGGTTATCCCTTGTGTCATGCTGTTACAGAGAGCTTGACAGTGCCA 429
QY 2481 TCTTAACGTCCTCTTTGGGGGAGAGCCACCTAACAGGAGGATTTTGTGAGGTGC 2540
Db 430 TTTTAACGTCCTTTGGGGGAGAGCCACCTAACAGGAGGATTTTGTGAGGTGC 489
QY 2541 CCTCTCTGAAAGAGTGGTGGGCAAGGCTTCTCTGGATCAAAATTCRAATAAATCAAG 2600
Db 490 CCTCTCTGAAAGAGTGGTGGGCAAGGCTTCTCTGGATCAAAATTCRAATAAATCAAG 549
QY 2601 TATTATTGATCTTAATATGTCAGGCTGGTCTAGAGCCAGCAGAGAAAGATTT 2660
Db 550 TATTATTGATCTTAATATGTCAGGCTGGTCTAGAGCCAGCAGAGAAAGATTT 609
QY 2661 ATACAGGACAGAGTCCCTTAACCTAAACATCCACAGGCCCCCAACTAGGAGTTTCAC 2720
Db 610 ATACAGGACAGAGTCCCTTAACCTAAACATCCACAGGCCCCCAACTAGGAGTTTCAC 669
QY 2721 TCATTCCAGTACTTTTAAAGCCGCTTGTGCTTTG 2758
Db 670 TCATTCCAGTACTTTTAAAGCCGCTTGTGCTTTG 707

RESULT 5
BB048324 637 bp mRNA linear EST 18-OCT-2001
LOCUS BB048324 RIKEN full-length enriched, adult male olfactory bulb Mus.
DEFINITION musculus cDNA clone 6430578P22 3', mRNA sequence.
"ESSTON BB048324
3ION BB048324.2 GI:16259477
SOURCE EST.
ORGANISM house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 637)
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiramoto,K., Hori,F., Ishi,Y., Ito,M., Kawai,J., Konno,H., Kouda
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okada
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
On Jun 11, 2000 this sequence version replaced gi:8455472.
Contact: Yoshinori Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
```

```

M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohata,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
FEATURES
Location/Qualifiers
1. 637
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="6430578P22"
/oligo_lib="RIKEN full-length enriched, adult male
olfactory bulb"
/sex="male"
/tissue_type="olfactory brain"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site 1: Sali; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGGATCCAGAGCTCTTTTCTTTTCTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 100.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGAGATTCGAGTAATAATTAATGCCCCCCCC
3']. cDNA was cloned into the XhoI and BamHI sites.
Vector: a modified pBluescript KS(+) after bulk excision
from Lambda FLC I. Cloning sites, 5' end: Sali; 3' end:
BamHI"
BASE COUNT 161 a 152 c 140 g 184 t
ORIGIN
Query Match 13.7%; Score 405; DB 10; Length 637;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 455; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2354 TGAAGTATACCCAGCTCTCCAGCCGACTGCAGATCAGGTTCCCTCCCTGATCTCTT 2413
Db 61 TGAAGTATACCCAGCTCTCTCCAGCCGACTGCAGATCAGGTTCCCTCCCTGATCTCTT 120
QY 2414 CTCATACCTGTGACCTCACAGGTTATCCCTTGTGCTCATGTGTTACAGAGAGCTTGA 2473
Db 121 CTCATACCTGTGACCTCACAGGTTATCCCTTGTGCTCATGTGTTACAGAGAGCTTGA 180
```

Qy	2474	GCTGCCATCTAAACGTGCTCTTTGGGGAGAGCCACCTAACAGGAGGATTTTGGTTTG	2533
Db	181	GCTGCCATCTAAACGTGCTCTTTGGGGAGAGCCACCTAACAGGAGGATTTTGGTTTG	240
Qy	2534	GAGTGCCCTCCTGAAAAAGTAGTGGGCAAAGGCTTCTCTGGGATCAAAATCAAAAT	2593
Db	241	GAGTGCCCTCCTGAAAAAGTAGTGGGCAAAGGCTTCTCTGGGATCAAAATCAAAAT	300
Qy	2594	AATCAAGTATTTTGAATGCTTAATATGTGCAAGGCGTGGTGCCTAGAGCCACGAGAA	2653
Db	301	AATCAAGTATTTTGAATGCTTAATATGTGCAAGGCGTGGTGCCTAGAGCCACGAGAA	360
Qy	2654	AGAATTTTATAACGAGCAGAGAAGTCCCTTAAACTTAAACATCCACAGGCCCCCAATCTAGGAG	2713
Db	361	AGAATTTTATAACGAGCAGAGAAGTCCCTTAAACTTAAACATCCACAGGCCCCCAATCTAGGAG	420
Qy	2714	GTTTCACTCCATCCAGTGACTTTTAAAGCGGCTTTGTGCGCTTGAATGCCCTTCTGA	2773
Db	421	GTTTCACTCCATCCAGTGACTTTTAAAGCGGCTTTGTGCGCTTGAATGCCCTTCTGA	480
Qy	2774	GATTTTGGATCTTCCTGTTCTCTGTCCCGCTGCTCCT	2809
...	481	GATTTTGGATCTTCCTGTTCTCTGTCCCGCTGCTCCT	516

RESULT	6
BB622895	
LOCUS	BB622895
DEFINITION	Riken full-length enriched, adult male olfactory brain Mus musculus cDNA clone 6430578r22 5', mRNA sequence.
ACCESSION	BB622895
VERSION	BB622895.1
KEYWORDS	GI:16461780
SOURCE	house mouse.
ORGANISM	Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
(bases 1 to 643)	
REFERENCE	Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai.K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa.A., Takahashi.F., Takeda.Y., Tanaka.T., Toyota.T., Muramatsu.M. and Hayashizaki.Y.
TITLE	Riken Mouse ESTs (AraKawa,T., et al. 2001)
JOURNAL	Unpublished (2001)

Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct. Func. Genomics 2 pre, L72-L86 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.
e mouse tissues.

FEATURES
source Location/Qualifiers
1. .670
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="5430403B13"
/clone_lib="RIKEN full-length enriched, 6 days neonate head"
/sex="mixed"
/tissue_type="head"
/dev_stage="6 days neonate"
/lab_host="DH10a"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGGATCCAGAGCTCTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCGAGTAAATAATATCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."

BASE COUNT 159 a 220 c 173 g 118 t
ORIGIN

Query Match 9.1%; Score 269; DB 10; Length 670;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 AGGAAGAGCTCACTATGGCTCCAGTCCCTCGGCATGCTGACTGCAGCCTGCAGCAAT 178
|||||
Db 251 AGGAAGAGCTCACTATGGCTCCAGTCCCTCGGCATGCTGACTGCAGCCTGCAGCAAT 310
|||||
Db 179 TTGGCGGCTCTAGCCCTCTCGGGGACTCAACACCCCTGGGGAAGAGGACACAAAGAGC 238
311 TTGGCGGCTCTAGCCCTCTCGGGGACTCAACACCCCTGGGGAAGAGGACACAAAGAGC 370
QY 239 CATACGCTGACCTTTTCAGCCCCAAAACCATGGGGGAGCCTACCCAGCTCCCTTCACAA 298
Db 371 CATACGCTGACCTTTTCAGCCCCAAAACCATGGGGGAGCCTACCCAGCTCCCTTCACAA 430
QY 299 GCACCAATGGACTCCTCTCTCTCGAGCGAGTCTCCGGGCCCGCCAGCCTCTGGCTATGCAA 358
Db 431 GCACCAATGGACTCCTCTCTCTCGAGCGAGTCTCCGGGCCCGCCAGCCTCTGGCTATGCAA 490
QY 359 ATGACTACCCACCCTTCCTCTCACTCACTATT 387
|||||
Db 491 ATGACTACCCACCCTTCCTCTCACTCACTATT 519

RESULT 10
AV337525 652 bp mRNA linear EST 31-AUG-2001
LOCUS
DEFINITION AV337525 RIKEN full-length enriched, adult male olfactory bulb Mus musculus cDNA clone 6430401K03 3', mRNA sequence.
ACCESSION AV337525
VERSION AV337525.2 GI:15406465
KEYWORDS EST.
SOURCE house mouse
ORGANISM Mus musculus

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-88L5"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."

BASE COUNT 49 a 33 c 39 g 43 t
ORIGIN

Query Match 3.4%; Score 101; DB 17; Length 164;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1997 CTTGAGTACTTAACCCCTATGCCCGAGGAGTAAGATAGTAATAATAATAATA 2056
|||||
Db 20 CTTGAGTACTTAACCCCTATGCCCGAGGAGTAAGATAGTAATAATAATAATA 79
|||||
QY 2057 ATTCTATTATCTAAGTTATGATGACGGGTACAGTACAGTG 2097
|||||
Db 80 ATTCTATTATCTAAGTTATGATGACGGGTACAGTACAGTG 120
|||||

Search completed: February 22, 2003, 02:41:26
Job time : 2693 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2003, 23:04:18 ; Search time 404 Seconds
(without alignments)
16499.790 Million cell updates/sec

Title: US-09-734-329-1
Perfect score: 2960
Sequence: 1 attctccattctccctccc.....aaaaaaaaaaaaaaaaaa 2960

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

arched: 2185239 seqs, 1125999159 residues

Word size : 0
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_101002:.*
1: /SID32/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID32/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID32/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID32/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID32/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID32/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID32/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID32/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID32/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID32/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID32/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID32/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID32/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID32/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID32/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID32/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID32/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID32/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID32/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID32/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID32/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID32/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID32/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID32/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	38	1.3	1025	20 AAX30319	DNA encoding a hum
C 2	38	1.3	1065	20 AAX30407	DNA encoding a hum
C 3	37	1.2	1396	24 ABA01972	Barley blotin synt
C 4	36	1.2	380	24 ABQ83319	Arabidopsis thalia
C 5	36	1.2	1049	21 AAF15600	Human prostate can
C 6	36	1.2	1095	20 AAX30366	DNA encoding a hum
C 7	36	1.2	1531	21 AAZ97040	Human secreted pro
C 8	35	1.2	52	18 AAX59252	Primer 4B used SIN
C 9	35	1.2	52	20 AAX58494	Primer 4B used SIN

C 10	35	1.2	65	18 AAX53261	Reverse primer SIN
C 11	35	1.2	65	20 AAX58503	Reverse primer SIN
C 12	35	1.2	108	24 ABL79250	Human ovarian canc
C 13	35	1.2	216	23 ABV57318	Human prostate exp
C 14	35	1.2	218	21 AAC98737	Human colon cancer
C 15	35	1.2	383	22 AAI88529	Human polynucleoti
C 16	35	1.2	387	24 ABQ83190	Arabidopsis thalia
C 17	35	1.2	396	22 AA191224	Human polynucleoti
C 18	35	1.2	422	22 AAI87358	Human polynucleoti
C 19	35	1.2	632	24 ABQ66100	Arabidopsis thalia
C 20	35	1.2	699	22 AAS33090	DNA encoding huma
C 21	35	1.2	767	24 ABN93163	Arabidopsis thalia
C 22	35	1.2	853	22 AAL01916	Human reproductive
C 23	35	1.2	853	23 ABL97209	Human testicular a
C 24	35	1.2	893	21 AAF65101	Membrane-bound pro
C 25	35	1.2	893	22 AAF44247	Human PRO1185 (UNQ
C 26	35	1.2	893	24 ABK40266	CDNA encoding huma
C 27	35	1.2	1231	21 AAZ52292	Maize replication
C 28	35	1.2	1234	21 AAZ34964	Soybean neutral tr
C 29	35	1.2	1354	21 AAC77921	Human cancer assoc
C 30	35	1.2	1399	21 AAF16022	Human prostate can
C 31	35	1.2	1972	24 ABK35347	Human CDNA encodin
C 32	35	1.2	1986	24 ABK84358	Human CDNA differe
C 33	35	1.2	2201	24 ABL90048	Human polynucleoti
C 34	35	1.2	2204	22 AAS26025	Human CDNA encodin
C 35	35	1.2	3288	22 AAH78736	Human H2LAD53 seri
C 36	35	1.2	5076	24 AAD29135	Venezuelan equine
C 37	35	1.2	5282	24 AAS61406	Human gene regulat
C 38	35	1.2	6240	24 ABL32136	Human immune syste
C 39	35	1.2	6240	24 ABL34454	Human metastasis a
C 40	35	1.2	6922	22 AAS46534	Tumour suppressor
C 41	35	1.2	6922	24 ABL70380	Chemically treated
C 42	35	1.2	6922	24 AAS61332	Human gene regulat
C 43	35	1.2	6922	24 ABK31411	Signal transductio
C 44	35	1.2	6989	24 AAD29137	Venezuelan equine
C 45	35	1.2	7970	20 AAX77357	Polynucleotide seq

ALIGNMENTS

102 (6) → not needed

RESULT 1
AAX30319/c
ID AAX30319 standard; DNA; 1025 BP.

XX AAX30319;

XX 14-MAY-1999 (first entry)

DE DNA encoding a human secreted protein.

XX Secreted protein; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; foetal deficiency; blood disorder;
KW CNS disorder; immune system disease; autoimmune disease; hepatic disease;
KW renal disease; diabetes; inflammation; allergy; ischemic shock;
KW Alzheimer's; cognitive disorder; schizophrenia; cardiovascular disorder;
KW prostate disease; asthma; osteoporosis; arthritis; ss.

OS Homo sapiens.

XX WO9907891-Al.

XX 18-FEB-1999.

XX 04-AUG-1998; 98WO-US16235.

XX 19-AUG-1997; 97US-0056732.

XX 05-AUG-1997; 97US-0054798.

XX 05-AUG-1997; 97US-0054803.

XX 05-AUG-1997; 97US-0054804.

XX 05-AUG-1997; 97US-0054806.

XX 05-AUG-1997; 97US-0054807.

XX 05-AUG-1997; 97US-0054808.

PR 05-AUG-1997; 97US-0054809.
PR 05-AUG-1997; 97US-0055309.
PR 05-AUG-1997; 97US-0055310.
PR 05-AUG-1997; 97US-0055312.
PR 05-AUG-1997; 97US-0055386.
PR 05-AUG-1997; 97US-0055311.
PR 18-AUG-1997; 97US-0055970.
PR 18-AUG-1997; 97US-0055986.
PR 19-AUG-1997; 97US-0056385.
PR 19-AUG-1997; 97US-0056386.
PR 19-AUG-1997; 97US-0056557.
PR 19-AUG-1997; 97US-0056370.
PR 19-AUG-1997; 97US-0056371.
PR 19-AUG-1997; 97US-0056563.
PR 19-AUG-1997; 97US-0056731.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Brewer LA, Ebner R, Ferrie AM, Greene JM, Janat F, Ni J;
PI Olsen HS, Rosen CA, Ruben SM, Soppet DR, Young PE, Yu G;
XX
XX WPI: 1999-167452/14.
DR P-PSDB; AAY10798.

PT New isolated human genes encoding secreted polypeptides - useful for
PT diagnosis and treatment of pathological diseases
PS Claim 3; Page 232; 331pp; English.
XX
XX The specification describes secreted proteins and their corresponding
CC polynucleotides which are useful for preventing, treating or ameliorating
CC medical conditions, e.g. by protein or gene therapy. Pathological
CC conditions can also be diagnosed by determining the presence of
CC secreted polypeptides in a sample or by determining the amount of the
CC mutations in the polynucleotides. Specific uses are described for each
CC of the products, based on which tissues they are most highly
CC expressed in, and include developing products for the diagnosis or
CC treatment of cancer, tumours, neurodegenerative disorders, developmental
CC abnormalities and foetal deficiencies, blood disorders, CNS disorders,
CC disease, diabetes, inflammation, allergies, ischemic shock, Alzheimer's
CC and cognitive disorders, schizophrenia, cardiovascular disorders,
CC prostate diseases, asthma, disorders involving osteoclasts such as
CC osteoporosis, arthritis or malignancies, diseases of testes, lung or
CC thymus, digestive/endocrine disorders, infections and AIDS. The
CC polypeptides are also useful for identifying their binding partners.
XX
SQ Sequence 1025 BP; 270 A; 251 C; 218 G; 286 T; 0 other;

Query Match 1.3%; Score 38; DB 20; Length 1025;
Best Local Similarity 100.0%; Pred. No. 6.8e-05;
atches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2131 AAGAAATCTAGTCAAAATGCATCTCTGTATAGACAAA 2168
|||||
Db 876 AAGAAATCTAGTCAAAATGCATCTCTGTATAGACAAA 839
|||||

RESULT 2
ID AAX30407
XX AAX30407 standard; DNA; 1065 BP.
XX AC
XX AAX30407;
XX
DT 14-MAY-1999 (first entry)
XX
DE DNA encoding a human secreted protein.
XX
XX Secreted protein; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; foetal deficiency; blood disorder;
KW CNS disorder; immune system disease; autoimmune disease; hepatic disease;
KW renal disease; diabetes; inflammation; allergy; ischemic shock;
KW Alzheimer's; cognitive disorder; schizophrenia; cardiovascular disorder;

KW prostate disease; asthma; osteoporosis; arthritis; ss.
XX Homo sapiens.
XX WO9907891-A1.
XX
XX 18-FEB-1999.
XX
PF 04-AUG-1998; 98WO-US16235.
XX
XX 19-AUG-1997; 97US-0056732.
XX 05-AUG-1997; 97US-0054798.
PR 05-AUG-1997; 97US-0054803.
PR 05-AUG-1997; 97US-0054804.
PR 05-AUG-1997; 97US-0054806.
PR 05-AUG-1997; 97US-0054807.
PR 05-AUG-1997; 97US-0054808.
PR 05-AUG-1997; 97US-0054809.
PR 05-AUG-1997; 97US-0055309.
PR 05-AUG-1997; 97US-0055310.
PR 05-AUG-1997; 97US-0055312.
PR 05-AUG-1997; 97US-0055386.
PR 05-AUG-1997; 97US-0055311.
PR 18-AUG-1997; 97US-0055970.
PR 18-AUG-1997; 97US-0055986.
PR 19-AUG-1997; 97US-0056385.
PR 19-AUG-1997; 97US-0056386.
PR 19-AUG-1997; 97US-0056557.
PR 19-AUG-1997; 97US-0056370.
PR 19-AUG-1997; 97US-0056371.
PR 19-AUG-1997; 97US-0056563.
PR 19-AUG-1997; 97US-0056731.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Brewer LA, Ebner R, Ferrie AM, Greene JM, Janat F, Ni J;
PI Olsen HS, Rosen CA, Ruben SM, Soppet DR, Young PE, Yu G;
XX
XX WPI: 1999-167452/14.
DR P-PSDB; AAY10886.

PT New isolated human genes encoding secreted polypeptides - useful for
PT diagnosis and treatment of pathological diseases
PS Claim 3; Page 270; 331pp; English.
XX
XX The specification describes secreted proteins and their corresponding
CC polynucleotides which are useful for preventing, treating or ameliorating
CC medical conditions, e.g. by protein or gene therapy. Pathological
CC conditions can also be diagnosed by determining the amount of the
CC secreted polypeptides in a sample or by determining the presence of
CC mutations in the polynucleotides. Specific uses are described for each
CC of the products, based on which tissues they are most highly
CC expressed in, and include developing products for the diagnosis or
CC treatment of cancer, tumours, neurodegenerative disorders, developmental
CC abnormalities and foetal deficiencies, blood disorders, CNS disorders,
CC disease, diabetes, inflammation, allergies, ischemic shock, Alzheimer's
CC and cognitive disorders, schizophrenia, cardiovascular disorders,
CC prostate diseases, asthma, disorders involving osteoclasts such as
CC osteoporosis, arthritis or malignancies, diseases of testes, lung or
CC thymus, digestive/endocrine disorders, infections and AIDS. The
CC polypeptides are also useful for identifying their binding partners.
XX
SQ Sequence 1065 BP; 292 A; 230 C; 263 G; 277 T; 3 other;

Query Match 1.3%; Score 38; DB 20; Length 1065;
Best Local Similarity 100.0%; Pred. No. 6.7e-05;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2131 AAGAAATCTAGTCAAAATGCATCTCTGTATAGACAAA 2168
|||||
Db 156 AAGAAATCTAGTCAAAATGCATCTCTGTATAGACAAA 193
|||||

```

RESULT 3
ABAO1972
ID ABA01972 standard; cDNA; 1396 BP.
XX
AC ABA01972;
XX
DT 06-FEB-2002 (first entry)
XX
DE Barley biotin synthase coding sequence #2.
XX
KW Barley; biotin synthase; biotin biosynthesis; herbicide; seedling;
XX plant development; clone bshl.pk0005.d10; transgenic plant; ss.
XX
OS Hordeum vulgare.
XX
Key Location/Qualifiers
CD5 3..1220
FT /*tag= a
FT /product= "biotin synthase"
FT /partial
FT /note= "the sequence has no start codon"
XX
PN US2001039042-A1.
XX
PD 08-NOV-2001.
XX
PF 19-DEC-2000; 2000US-0740288.
XX
PR 21-DEC-1999; 99US-172929P.
XX
PA (ALLE/) ALLEN S M.
PA (KINN/) KINNEY A J.
PA (MIAO/) MIAO G.
PA (OROZ/) OROZCO E M.
XX
PI Allen SM, Kinney AJ, Miao G, Orozco EM;
XX
DR WPI; 2002-040723/05.
XX
P-PSDB; AAM51979.
XX
PT New polypeptides, useful as targets for herbicide discovery, and as
ET probes for genetic and physical mapping of genes of which they are
part, or creating transgenic plants, comprises biotin synthase
polypeptides and encoding polynucleotides.
PS Claim 5; Page 24; 46pp; English.
XX
CC The present invention provides the protein and coding sequences of biotin
CC synthase enzymes from barley, wheat, maize, soybean and the prickly
CC poppy. These sequences can be used to produce transgenic plants which
CC express different levels of the gene, or express it at different times in
CC plant development. They can also be used as a target in the production of
CC herbicides. The present sequence is a barley biotin synthase cDNA
CC obtained from clone bshl.pk0005.d10, which was derived from a barley
CC seedling.
XX
SQ Sequence 1396 BP; 387 A; 351 C; 356 G; 302 T; 0 other;
XX
Query Match 1.2%; Score 37; DB 24; Length 1396;
Best Local Similarity 100.0%; Pred. No. 0.00015;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 2924 CGATTTCACAAAAA...AAAAAAAAAAAAAAAAAAAAA 2960
|||||
DB 1349 CGATTTCACAAAAA...AAAAAAAAAAAAAAAAAAAAA 1385
|||||
RESULT 4
ABQ85319/c
ID ABQ85319 standard; DNA; 380 BP.
XX
AC ABQ85319;
XX
DT 05-SEP-2002 (first entry)
XX
DE Arabidopsis thaliana expressed polynucleotide SEQ ID NO 189.
XX
KW Plant; Arabidopsis; transgenic; fungicide; insecticide; ds.
XX
OS Arabidopsis thaliana.
XX
PN US2002062014-A1.
XX
PD 23-MAY-2002.
XX
PF 26-JAN-2001; 2001US-0770791.
XX
PR 27-JAN-2000; 2000US-178480P.
XX
(GORL/) GORLACH J.
XX
(PANY/) AN Y.
XX
(PHAMI/) HAMILTON C M.
XX
(PRIC/) PRICE J L.
XX
(PRAIN/) RAINES T M.
XX
(PYUY/) YU Y.
XX
(PAME/) RAMEAKA J G.
XX
(PAGE/) PAGE A.
XX
(PMATH/) MATHEW A V.
XX
(PLEDE/) LEDFORD B L.
XX
(PWOES/) WOESSNER J P.
XX
(PHAAS/) HAAS W D.
XX
(PGARC/) GARCIA C A.
XX
(PKRIC/) KRICKER M.
XX
(PSLAT/) SLATER T.
XX
(PDAVI/) DAVIS K R.
XX
(PALLE/) ALLEN K.
XX
(PHOFF/) HOFFMAN N.
XX
(PHURE/) HUREAN P.
XX
PI Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y;
XX
P Rameaka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;
XX
P Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;
XX
P Hurban P;
XX
WPI; 2002-479265/51.
XX
PT New nucleic acid sequences of Arabidopsis thaliana and their encoded
PT products are useful to produce transgenic plants, to screen for
PT biologically active agents such as fungicides and insecticides and in
PT genetic studies.
XX
PS Claim 1; SEQ ID NO 189; 18pp + Sequence Listing; English.
XX
CC The invention relates to a novel nucleic acid of Arabidopsis thaliana
CC comprising a sequence capable of hybridizing under stringency to one of
CC the 999 sequences referred to but not defined in the specification
CC (ABQ85131-ABQ86129). The nucleic acid sequences are useful to identify
CC homologous or related genes, to produce compositions that modulate
CC expression or function of the encoded protein, to map functional regions
CC of the protein, to study associated physiological pathways, to
CC genetically manipulate cells and plants. The encoded products are useful
CC to screen for biologically active agents such as fungicides or
CC insecticides and to elucidate biochemical pathways.
XX
SQ Sequence 380 BP; 96 A; 74 C; 61 G; 149 T; 0 other;
XX
Query Match 1.2%; Score 36; DB 24; Length 380;
Best Local Similarity 100.0%; Pred. No. 0.00045;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 2925 GATTTCACAAAAA...AAAAAAAAAAAAAAAAAAAAA 2960
|||||
DB 69 GATTTCACAAAAA...AAAAAAAAAAAAAAAAAAAAA 34
|||||

```

RESULT 5
 AAF15600
 ID AAF15600 standard; cDNA; 1049 BP.
 XX
 AC AAF15600;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Human prostate cancer antigen nucleotide sequence SEQ ID NO:35.
 XX
 KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
 KW neuroprotective; cytotactic; cardioactive; immunomodulatory; muscular;
 KW vulnary; gastrointestinal; nephrotropic; antineoplastic; gynaecological;
 KW antibacterial; gene therapy; neural; immune; reproductive; renal;
 KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
 KW wound; infectious disease; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO2000055174-A1.
 XX
 PD 18-FEB-1999.
 XX
 PF 04-AUG-1998; 98WO-US16235.
 XX
 PR 19-AUG-1997; 97US-0056732.
 PR 05-AUG-1997; 97US-0054798.
 PR 05-AUG-1997; 97US-0054803.
 PR 05-AUG-1997; 97US-0054804.
 PR 05-AUG-1997; 97US-0054806.
 PR 05-AUG-1997; 97US-0054807.
 PR 05-AUG-1997; 97US-0054808.
 PR 05-AUG-1997; 97US-0054809.
 PR 05-AUG-1997; 97US-0055310.
 PR 05-AUG-1997; 97US-0055312.
 PR 05-AUG-1997; 97US-0055386.
 PR 05-AUG-1997; 97US-0055311.
 PR 18-AUG-1997; 97US-0055970.
 PR 19-AUG-1997; 97US-0055986.
 PR 19-AUG-1997; 97US-0056365.
 PR 19-AUG-1997; 97US-0056366.
 PR 19-AUG-1997; 97US-0056557.
 PR 19-AUG-1997; 97US-0056370.
 PR 19-AUG-1997; 97US-0056371.
 PR 19-AUG-1997; 97US-0056563.
 PR 19-AUG-1997; 97US-0056731.
 XX
 (HUMA-) HUMAN GENOME SCI INC.
 XX
 PA Brewer LA, Ebner R, Ferrie AM, Greene JM, Janat F, Ni J;
 PI Olsen HS, Rosen CA, Ruben SM, Soppet DR, Young PE, Yu G;
 XX
 DR WPI: 1999-167452/14.
 DR P-PSDB; AAY10845.
 XX
 PT New isolated human genes encoding secreted polypeptides - useful for
 PT diagnosis and treatment of pathological diseases
 XX
 PS Claim 3; Page 253; 331pp; English.
 XX
 CC The specification describes secreted proteins and their corresponding
 CC polynucleotides which are useful for preventing, treating or ameliorating
 CC medical conditions, e.g. by protein or gene therapy. Pathological
 CC conditions can also be diagnosed by determining the amount of the
 CC secreted polypeptides in a sample or by determining the presence of
 CC mutations in the polynucleotides. Specific uses are described for each
 CC of the products, based on which tissues they are most highly
 CC expressed in, and include developing products for the diagnosis or
 CC treatment of cancer, tumours, neurodegenerative disorders, developmental
 CC abnormalities and foetal deficiencies, blood disorders, CNS disorders,
 CC diseases of the immune system, autoimmune diseases, hepatic and renal,
 CC disease, diabetes, inflammation, allergies, ischemic shock, Alzheimer's
 CC and cognitive disorders, schizophrenia, cardiovascular disorders,
 CC prostate diseases, asthma, disorders involving osteoclasts such as
 CC osteoporosis, arthritis or malignancies, diseases of testes, lung or
 CC thymus, digestive/endocrine disorders, infections and AIDS. The
 CC polypeptides are also useful for identifying their binding partners.

Query Match 1.2%; Score 36; DB 21; Length 1049;
 Best Local Similarity 100.0%; Pred. No. 0.00039;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2925 GATTTCACAAAAA...AAAAAAAAAAAAAAAAAAAA 2960
 Db 995 GATTTCACAAAAA...AAAAAAAAAAAAAAAAAAAA 1030
 RESULT 6
 AAX30366
 ID AAX30366 standard; DNA; 1095 BP.
 XX
 AC AAX30366;

```
XX SQ Sequence 1095 BP; 225 A; 365 C; 282 G; 219 T; 4 other;
Query Match 1.2%; Score 36; DB 20; Length 1095;
Best Local Similarity 100.0%; Pred. No. 0.00038;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2925 GATTTCACAAAAA...AAAAAAAAAAAAAAAAAAAAA 2960
|||||
Db 1044 GATTTCACAAAAA...AAAAAAAAAAAAAAAAAAAAA 1079

RESULT 7
AAZ97040
ID AAZ97040 standard; cDNA; 1531 BP.
XX AAZ97040;

19-APR-2000 (first entry)
Human secreted protein gene 22 cDNA clone HOHCK70, SEQ ID NO:32.
Human; secreted protein; cancer; tumour; developmental abnormality;
foetal deficiency; blood disorder; immune system disorder; inflammation;
autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;
schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;
atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;
digestive disorder; endocrine disorder; infection; AIDS; leukaemia;
therapy; ds.
XX Homo sapiens.
XX WO9966041-A1.
XX 23-DEC-1999.
XX 15-JUN-1999; 99WO-US13418.
XX 16-JUN-1998; 98US-0089507.
PR 16-JUN-1998; 98US-0089508.
PR 16-JUN-1998; 98US-0089509.
PR 16-JUN-1998; 98US-0089510.
PR 22-JUN-1998; 98US-0090112.
PR 22-JUN-1998; 98US-0090113.
XX (HUMA-) HUMAN GENOME SCI INC.

PI Ruben SM, Ni J, Rosen CA, Wei Y, Young PE, Florence KA;
PI Soppet DR, Brewer LA, Endress GA, Carter KC, Mucenski M, Ebner R;
PI Lafleur DW, Olsen HS, Shi Y, Moore PA, Komatsoulis G;
XX
XX WPI; 2000-106100/09.
XX P-PSDB; AAY86236.
XX New isolated human genes and the secreted polypeptides they encode,
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders -
XX
XX Claim 1; Page 330; 586pp; English.
XX
XX AAZ97019 to AAZ97137 represent 94 isolated human secreted protein genes.
CC AAZ96215 to AAZ86333 are the secreted proteins encoded by the 94 human
CC genes. This sequence represents a fragment of one of the human secreted
CC proteins. The genes and their corresponding secreted polypeptides are
CC useful for preventing, treating or ameliorating medical conditions,
CC e.g., by protein or gene therapy. Also pathological conditions can be
CC diagnosed by determining the amount of the new polypeptides in a sample
CC or by determining the presence of mutations in the new genes. Specific
CC uses are described for each of the 94 genes, based on which tissues they
CC are most highly expressed in, and include developing products for the
CC diagnosis or treatment of cancer, tumours, developmental abnormalities
CC and foetal deficiencies, blood disorders, diseases of the immune system,
CC autoimmune diseases, inflammation, allergies, Alzheimer's and cognitive
```

```
CC disorders, schizophrenia, arthritis, asthma, psoriasis, sepsis, skin
CC disorders, atherosclerosis, diabetes, cardiovascular disorders, kidney
CC disorders, digestive/endocrine disorders, infections and AIDS. The
CC polypeptides are also useful for identifying their binding partners.
CC The sequences shown in AAY86334 to AAY86585 represent fragments of the
CC secreted proteins.
XX
SQ Sequence 1531 BP; 479 A; 285 C; 268 G; 499 T; 0 other;
Query Match 1.2%; Score 36; DB 21; Length 1531;
Best Local Similarity 100.0%; Pred. No. 0.00036;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2925 GATTTCACAAAAA...AAAAAAAAAAAAAAAAAAAAA 2960
|||||
Db 1488 GATTTCACAAAAA...AAAAAAAAAAAAAAAAAAAAA 1523

RESULT 8
AAZ59252/c
ID AAZ59252 standard; DNA; 52 BP.
XX AC AAZ59252;
XX 06-SEP-1999 (first entry)
DT Primer 4B used SIN-1 cDNA PCR.
DE SIN-1; alphavirus; vector; infection; cancer; autoimmune disease;
KW gene therapy; vaccine; primer; PCR; ss.
KW Synthetic.
OS Sindbis virus.
XX WO9738087-A2.
XX 16-OCT-1997.
XX 04-APR-1997; 97WO-US06010.
XX 12-JUL-1996; 96US-0679640.
PR 05-APR-1996; 96US-0628594.
PR 24-JUN-1996; 96US-0668953.
XX (CHIR ) CHIRON VIAGENE INC.
XX (UNIW ) UNIV WASHINGTON.
XX
XX Belli BA, Dryga SA, Dubensky TW, Frolov I, Polo JM;
PI Schlesinger S;
XX WPI; 1997-512707/47.
XX
XX Nucleic acid comprising altered alpha-virus non-structural protein
PT gene - useful for generating expression cassettes for production of
PT recombinant proteins in vertebrate or insect cells
XX
XX Example 1; Page 99; 309pp; English.
XX
XX This primer oligonucleotide, termed 4B, includes an XbaI site, a
CC T35 sequence, and nucleotides 11703-11698 of the Sindbis virus
CC genome. It was used in the PCR amplification of SIN-1 cDNA. A set
CC of primer pairs necessary for amplifying the entire SIN-1 genome is
CC provided (see AAX59241-52). SIN-1 (see AAX59321) is a Sindbis virus
CC variant strain which exhibits reduced inhibition of host
CC macromolecular synthesis and which is capable of establishing
CC persistent infection in vertebrate cells, in contrast to lytic,
CC cytopathogenic wild-type strains of the same virus. The invention
CC relates to alphavirus-based vectors with reduced inhibition of
CC cellular macromolecular synthesis. Alphavirus vector constructs,
CC replicons and eukaryotic layered vector initiation systems are used:
CC (1) to deliver a selected heterologous sequence, particularly in
CC gene therapy for treatment of a wide range of infections, cancers,
CC and autoimmune diseases, or to regulate the immune system; (11) as
```

CC vaccines: (iii) to inhibit pathogens; and (iv) to express
 CC heterologous products (therapeutic proteins, ribozymes, and
 CC antisense sequences). Since the modified vectors do not cause
 CC significant inhibition of host cell biosynthesis, they can be used
 CC safely as gene therapy vectors.

XX SQ Sequence 52 BP; 7 A; 1 C; 3 G; 41 T; 0 other;

Query Match 1.2%; Score 35; DB 18; Length 52;
 Best Local Similarity 100.0%; Pred. No. 0.0015;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2926 ATTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2960
 |||||
 Db 51 ATTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 17

RESULT 9
 AAX58494/C
 ID AAX58494 standard; DNA; 52 BP.
 XX
 AC AAX58494;

16-AUG-1999 (first entry)
 DE Primer 4B used SIN-1 CDNA PCR.

XX SIN-1; alphavirus; vector; infection; cancer; autoimmune disease;
 KW gene therapy; vaccine; primer; PCR; ss.

XX Synthetic.
 OS Sindbis virus.
 XX WO9918226-A2.

XX 15-APR-1999.
 XX 06-OCT-1998; 98WO-US21062.
 XX 06-OCT-1997; 97US-0944465.

XX (CHIR) CHIRON CORP.
 XX (UNIW) UNIV WASHINGTON.

XX Belli BA, Dryga SA, Dubensky TW, Frolov I, Polo JM;
 XX Schlesinger S;

XX WPI; 1999-264032/22.

XX Alphavirus vectors with reduced cytopathic effects

XX Example 1; Page 103; 235pp; English.

XX This primer oligonucleotide, termed 4B, includes an XbaI site, a
 CC T35 sequence, and nucleotides 11703-11698 of the Sindbis virus
 CC genome. It was used in the PCR amplification of SIN-1 cDNA. A set
 CC of primer pairs necessary for amplifying the entire SIN-1 genome is
 CC provided (see AAX58483-94). SIN-1 (see AAX58571) is a Sindbis virus
 CC variant strain which exhibits reduced inhibition of host
 CC macromolecular synthesis and which is capable of establishing
 CC persistent infection in vertebrate cells, in contrast to lytic,
 CC cytopathogenic wild-type strains of the same virus. The invention
 CC relates to alphavirus-based vectors with reduced inhibition of
 CC cellular macromolecular synthesis. Alphavirus vector constructs,
 CC replicons and eukaryotic layered vector initiation systems are used:
 CC (i) to deliver a selected heterologous sequence, particularly in
 CC gene therapy for treatment of a wide range of infections, cancers,
 CC and autoimmune diseases, or to regulate the immune system; (ii) as
 CC vaccines; (iii) to inhibit pathogens; and (iv) to express
 CC heterologous products (therapeutic proteins, ribozymes, and
 CC antisense sequences). Since the modified vectors do not cause
 CC significant inhibition of host cell biosynthesis, they can be used
 CC safely as gene therapy vectors.

XX SQ Sequence 52 BP; 7 A; 1 C; 3 G; 41 T; 0 other;

Query Match 1.2%; Score 35; DB 20; Length 52;
 Best Local Similarity 100.0%; Pred. No. 0.0015;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2926 ATTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2960
 |||||
 Db 51 ATTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 17

RESULT 10
 AAX59261/C
 ID AAX59261 standard; DNA; 65 BP.

XX AAX59261;

XX 06-SEP-1999 (first entry)

XX Reverse primer SIN11703R used in SIN-1 vector construction.

XX SIN-1; alphavirus; vector; infection; cancer; autoimmune disease;
 KW gene therapy; vaccine; primer; PCR; ss.

XX Synthetic.
 OS Sindbis virus.
 XX WO9738087-A2.

XX 16-OCT-1997.

XX 04-APR-1997; 97WO-US06010.

XX 12-JUL-1996; 96US-0679640.

XX 05-APR-1996; 96US-0628594.

XX 24-JUN-1996; 96US-0668953.

XX (CHIR) CHIRON VIAGENE INC.
 XX (UNIW) UNIV WASHINGTON.

XX Belli BA, Dryga SA, Dubensky TW, Frolov I, Polo JM;
 XX Schlesinger S;

XX WPI; 1997-512707/47.

XX Nucleic acid comprising altered alpha-virus non-structural protein
 CC gene - useful for generating expression cassettes for production of
 CC recombinant proteins in vertebrate or insect cells

XX Example 3; Page 119; 309pp; English.

XX Reverse primer SIN11703R corresponds to nucleotides 11703-11698 of
 CC the Sindbis virus genome and also includes a buffer sequence, SacI
 CC and PmeI restriction sites and a T40 tract. It was used with
 CC primer SIN11386F (see AAX59260), in the PCR amplification of the 3'
 CC region of a SIN-1 derived vector comprising the viral 3' end, a
 CC polyA tract and unique restriction recognition sequence. SIN-1-based
 CC RNA vector replicons were constructed. SIN-1 (see AAX59321) is a
 CC Sindbis virus variant which exhibits reduced inhibition of host
 CC macromolecular synthesis and is capable of establishing persistent
 CC infection in vertebrate cells, in contrast to lytic, cytopathogenic
 CC wild-type strains of the virus. The invention relates to alphavirus
 CC (e.g. Sindbis virus)-based vectors with reduced inhibition of
 CC cellular macromolecular synthesis. Alphavirus vector constructs,
 CC replicons and eukaryotic layered vector initiation systems are used:
 CC (i) to deliver a selected heterologous sequence, particularly in
 CC gene therapy for treatment of a wide range of infections, cancers,
 CC and autoimmune diseases, or to regulate the immune system; (ii) as
 CC vaccines; (iii) to inhibit pathogens; and (iv) to express
 CC heterologous products (therapeutic proteins, ribozymes, and
 CC antisense sequences). Since the modified vectors do not cause
 CC significant inhibition of host cell biosynthesis, they can be used

CC safely as gene therapy vectors.

XX Sequence 65 BP; 9 A; 4 C; 5 G; 47 T; 0 other;

SO Query Match 1.2%; Score 35; DB 18; Length 65;

Best Local Similarity 100.0%; Pred. No. 0.0014;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2926 ATTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2960

|||||

Db 64 ATTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 30

RESULT 11

ABL79250/C

ID ABL79250 standard; CDNA; 108 BP.

XX

AC ABL79250;

DT 17-MAY-2002 (first entry)

XX Human ovarian cancer related CDNA clone SEQ ID NO:2228.

DE Reverse primer SIN11703R used in SIN-1 vector construction.

XX SIN-1; alphavirus; vector; infection; cancer; autoimmune disease;

KW gene therapy; vaccine; primer; PCR; ss.

XX Synthetic.

OS Sindbis virus.

XX WO9918226-A2.

PN 15-APR-1999.

XX 06-OCT-1998; 98WO-US21062.

XX 06-OCT-1997; 97US-0944465.

XX (CHIR) CHIRON CORP.

PA (UNIW) UNIV WASHINGTON.

XX Belli BA, Dryga SA, Dubensky TW, Frolov I, Polo JM;

PI Schlesinger S;

XX WPI: 1999-264032/22.

DR Alphavirus vectors with reduced cytopathic effects

XX Example 3; Page 124; 235pp; English.

CC Reverse primer SIN11703R corresponds to nucleotides 11703-11698 of

CC the Sindbis virus genome and also includes a buffer sequence, SacI

CC and PmeI restriction sites and a T40 tract. It was used with

CC primer SIN11386F (see AAX58502), in the PCR amplification of the 3'

CC region of a SIN-1 derived vector comprising the viral 3' end, a

CC polyA tract and unique restriction recognition sequence. SIN-1-based

CC RNA vector replicons were constructed. SIN-1 (see AAX58571) is a

CC Sindbis virus variant which exhibits reduced inhibition of host

CC macromolecular synthesis and is capable of establishing persistent

CC infection in vertebrate cells, in contrast to lytic, cytopathogenic

CC wild-type strains of the virus. The invention relates to alphavirus

CC (e.g. Sindbis virus)-based vectors with reduced inhibition of

CC cellular macromolecular synthesis. Alphavirus vector constructs,

CC replicons and eukaryotic layered vector initiation systems are used:

CC (i) to deliver a selected heterologous sequence, particularly in

CC gene therapy for treatment of a wide range of infections, cancers,

CC and autoimmune diseases, or to regulate the immune system; (ii) as

CC vaccines; (iii) to inhibit pathogens; and (iv) to express

CC heterologous products (therapeutic proteins, ribozymes, and

CC antisense sequences). Since the modified vectors do not cause

CC significant inhibition of host cell biosynthesis, they can be used

CC safely as gene therapy vectors.

XX Sequence 65 BP; 9 A; 4 C; 5 G; 47 T; 0 other;

Query Match

Best Local Similarity 100.0%; Pred. No. 0.0014;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2926 ATTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2960

|||||

Db 64 ATTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 30

RESULT 12

ABL79250/C

ID ABL79250 standard; CDNA; 108 BP.

XX

AC ABL79250;

DT 17-MAY-2002 (first entry)

XX Human ovarian cancer related CDNA clone SEQ ID NO:2228.

DE Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.

KW Homo sapiens.

XX WO200192581-A2.

PN 06-DEC-2001.

XX 29-MAY-2001; 2001WO-US17756.

XX 26-MAY-2000; 2000US-207484P.

XX (CORI-) CORIXA CORP.

PA Algate PA, Harlocker SL, Jones R;

XX WPI: 2002-122075/16.

DR Composition for therapy and diagnosis of ovarian cancer comprising

XX polypeptide of a ovarian tumor polypeptide, polynucleotide encoding

XX polypeptide, antibody specific to polypeptide or T cell expressing

XX polypeptide

XX Claim 1; SEQ ID 2228; 489pp; English.

XX The present invention describes a composition (I) comprising: carriers

XX and immunostimulants; and a polypeptide (II) of a ovarian tumour

XX polypeptide encoded by a polynucleotide (III) having a cDNA sequence

XX (SI) from the 10912 nucleotide sequences as given in ABL77023 to

XX ABL87934, (III) encoding (II) having a sequence (S2), a T cell

XX population of (II), or antigen presenting cells that express (II).

XX (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to

XX (SI) can be used for detecting ovarian cancer in a patient's biological

XX sample preferably serum or ovarian tissue. The method comprises

XX contacting a biological sample from a patient with (IV), detecting the

XX amount of polynucleotide hybridising to (IV) and comparing the amount to

XX a predetermined cutoff value and thereby detecting ovarian cancer in the

XX patient, where the amount of polynucleotide hybridising to (IV) is

XX detected preferably by polymerase chain reaction (PCR). (I) comprising

XX (III) and/or (II) is useful for stimulating and/or expanding T cells

XX specific for an ovarian tumour protein comprising contacting T cells

XX with (III) or (II). (III) is useful in design and preparation of

XX CC ribozyme molecules for inhibiting expression of the tumour polypeptides

XX and proteins in tumour cells; and to isolate a full length gene from a

XX suitable library e.g., a tumour cDNA library using well known

XX techniques.

XX Sequence 108 BP; 18 A; 20 C; 15 G; 55 T; 0 other;

Query Match

Best Local Similarity 100.0%; Pred. No. 0.0013;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2926 ATTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2960
DB 35 ATTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1

RESULT 13

ABV57318
ID ABV57318 standard; cDNA; 216 BP.

AC ABV57318;
XX
XX
XX
XX 17-SEP-2002 (first entry)
XX Human prostate expression marker cDNA 57309.
XX
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
XX Homo sapiens.

OS
XX
XX WO200160860-A2.
PN
XX 23-AUG-2001.

PF 20-FEB-2001; 2001WO-US05171.

XX 17-FEB-2000; 2000US-183319P.

PR 16-MAR-2000; 2000US-189862P.

PR 25-MAY-2000; 2000US-207454P.

PR 09-JUN-2000; 2000US-211314P.

PR 18-JUL-2000; 2000US-219007P.

PR 13-DEC-2000; 2000US-255281P.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Schlegel R, Endege WO, Monahan JE;

PI
XX WPI; 2001-662795/76.

XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -

PS Claim 1; Page 11028; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;

CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;

CC (e) selecting a composition for inhibiting prostate cancer in a patient;

CC (f) assessing the prostate cell carcinogenic potential of a compound;

CC (g) determining whether prostate cancer has metastasized in a patient;

CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;

CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX Sequence 216 BP; 104 A; 33 C; 42 G; 37 T; 0 other;

Query Match 1.2%; Score 35; DB 23; Length 216;

Best Local Similarity 100.0%; Pred. No. 0.0012;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2926 ATTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2960

DB 143 ATTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 177

RESULT 14

AAC98737
ID AAC98737 standard; cDNA; 218 BP.

XX AAC98737;

XX
XX
XX 09-MAR-2001 (first entry)

XX Human colon cancer antigen nucleotide sequence SEQ ID NO:747.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW identification; cytostatic; cardioactive; neuroprotective; vulnary;
KW immunomodulatory; muscular; gynaecological; gastrointestinal;
KW nephrotropic; anti-infective; antibacterial; gene therapy; wound;
KW neural disorder; immune system disorder; muscular disorder;
KW reproductive disorder; gastrointestinal disorder; renal disorder;
KW infectious disease; cardiovascular disorder; ss.

XX Homo sapiens.

XX WO200055351-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05883.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI; 2000-587534/55.

XX P-PSDB; AAB53980.

XX Colon cancer associated gene sequences, referred to as colon cancer
PT antigens, useful for the treatment, prevention, and diagnosis of colon
PT disorders such as colon cancer -

PS Claim 1; Page 1306; 2104pp; English.

XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC vulnary, nephrotropic, anti-infective and antibacterial activities, and
CC can be used in gene therapy. The colon cancer antigen polynucleotides,
CC proteins and antibodies to the proteins are useful for the prevention,
CC treatment and diagnosis of colon disorders, such as colon cancer. The
CC polynucleotides may be used in diagnostics and research, such as for
CC chromosome identification, and as hybridisation probes. The proteins
CC may also be used to prevent diseases such as neural disorders, immune
CC system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, wounds, renal disorders, infectious
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC AAB54007 represent sequences used in the exemplification of the present
CC invention.

XX Sequence 218 BP; 115 A; 22 C; 26 G; 50 T; 5 other;

Query Match 1.2%; Score 35; DB 21; Length 218;

Best Local Similarity 100.0%; Pred. No. 0.0012;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2926 ATTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2960

DB 98 ATTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 132

RESULT 15

AAB8529

ID AAB8529 standard; cDNA; 383 BP.

XX
XX AC AAB8529;

XX	06-NOV-2001	(first entry)
DT	Human polynucleotide SEQ ID NO 8589.	
XX		
DE	Human; cytokine; cell proliferation; cell differentiation; gene therapy;	
XX	vaccine; peptide therapy; stem cell growth factor; haematopoiesis;	
KW	tissue growth factor; immunomodulatory; cancer; leukaemia;	
KW	nervous system disorders; arthritis; inflammation; ss.	
XX		
XX	Homo sapiens.	
OS		
XX	WO2001164835-A2.	
PN		
XX		
XX	07-SEP-2001.	
PD		
PX		
PF	26-FEB-2001; 2001WO-US04927.	
PR	28-FEB-2000; 2000US-0515126.	
XX	18-MAY-2000; 2000US-0577409.	
PA	(HYSE-) HYSEQ INC.	
PI	Tang YT, Liu C, Drmanac RT;	
XX		
DR	WPI; 2001-514838/56.	
DR	P-PSDB; AAO08598.	
XX		
PT	Isolated nucleic acids and polypeptides, useful for preventing	
PT	diagnosing and treating e.g. leukaemia, inflammation and immune	
PT	disorders -	
XX		
XX	Claim 1: SEQ ID NO 8589; 1399pp + Sequence Listing; English.	
PS		
CC	The invention relates to human polynucleotides (AAI79941-AAI93841) and	
CC	the encoded proteins (AAO0010-AAO13910) that exhibit activity elating to	
CC	'cytokine, cell proliferation or cell differentiation or which may induce	
CC	production of other cytokines in other cell populations. The	
CC	polynucleotides and polypeptides are useful in gene therapy, vaccines or	
CC	peptide therapy. The polypeptides have various cytokine-like activities,	
CC	e.g. stem cell growth factor activity, haematopoiesis regulating	
CC	activity, tissue growth factor activity, immunomodulatory activity and	
CC	activin/inhibin activity and may be useful in the diagnosis and/or	
CC	treatment of cancer, leukaemia, nervous system disorders, arthritis and	
CC	inflammation.	
CC	Note: The sequence data for this patent did not form part of the printed	
	specification, but was obtained in electronic format directly from WIPO	
	at ftp.wipo.int/pub/published_pct_sequences.	
XX		
SQ	Sequence 383 BP; 162 A; 41 C; 60 G; 120 T; 0 other;	
	Query Match 1.2%; Score 35; DB 22; Length 383;	
	Best Local Similarity 100.0%; Pred. No. 0.0011;	
	Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	2926 ATTCTCAAAAAAAAAAAAAAAAAAAAAA 2960	
Db	186 ATTTCAAAAAAAAAAAAAAAAAAAAAA 220	

Search completed: February 22, 2003, 00:31:46
Job time : 424 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: February 21, 2003, 23:05:53 ; Search time 5064 Seconds
(without alignments)
17011.119 Million cell updates/sec

Title: US-09-734-329-1
Perfect score: 2960
Sequence: 1 attctccattctccctccc.....aaaaaaaaaaaaaaaaaaaaa 2960

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

arched: 2054640 seqs, 14551402878 residues

Word size : 0
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg_hum.*
- 31: em.htg_inv.*
- 32: em.htg_other.*
- 33: em.htg_mus.*
- 34: em.htg_pln.*
- 35: em.htg_rod.*
- 36: em.htg_mam.*
- 37: em.htg_vrt.*
- 38: em.sy.*
- 39: em.htgo_hum.*
- 40: em.htgo_mus.*
- 41: em.htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2960	100.0	2960	10	AF184902	AF184902 Mus muscu
2	1190	40.2	138860	2	AC055703	AC055703 Mus muscu
3	121	4.1	138860	2	AC055703	AC055703 Mus muscu
4	55	1.9	2395	9	AF477981	AF477981 Homo sapi
5	55	1.9	166697	2	AC021103	AC021103 Homo sapi
6	55	1.9	192665	2	AC073611	AC073611 Homo sapi
7	36	1.2	819	8	LAU89841	U89841 Lupinus ang
8	36	1.2	1451	3	AF201906	AF201906 Drosophil
9	36	1.2	1683	9	BC015812	BC015812 Homo sapi
10	36	1.2	2108	9	BC010743	BC010743 Homo sapi
11	36	1.2	2185	10	BC005633	BC005633 Mus muscu
12	36	1.2	2209	9	BC002471	BC002471 Homo sapi
13	36	1.2	2255	3	AF201905	AF201905 Drosophil
14	36	1.2	2455	5	AF022890	AF022890 Gallus ga
15	36	1.2	93491	2	AC116967	AC116967 Dictyoste
16	36	1.2	231574	2	AC125119	AC125119 Mus muscu
17	35	1.2	478	3	AF352722	AF352722 Cooperia
18	35	1.2	542	8	AF336985	AF336985 Vaucheria
19	35	1.2	654	5	AY124337	AY124337 Plecoglos
20	35	1.2	790	3	AY113627	AY113627 Drosophil
21	35	1.2	837	8	AF015785	AF015785 Phaseolus
22	35	1.2	854	9	BC005921	BC005921 Homo sapi
23	35	1.2	866	3	AY118616	AY118616 Drosophil
24	35	1.2	893	6	AX201346	AX201346 Sequence
25	35	1.2	893	6	AX403513	AX403513 Sequence
26	35	1.2	902	9	BC014348	BC014348 Homo sapi
27	35	1.2	907	10	CDPPTKI	Z50782 C.porceillus
28	35	1.2	961	10	CPGPPTKI	Z50783 C.porceillus
29	35	1.2	1006	10	CPBPPTKI	Z50784 C.porceillus
30	35	1.2	1044	9	BC016306	BC016306 Homo sapi
31	35	1.2	1056	10	AF038848	AF038848 Mus muscu
32	35	1.2	1210	9	BC009956	BC009956 Homo sapi
33	35	1.2	1239	10	AF000577	AF000577 Rattus no
34	35	1.2	1246	9	HSV526F1B	270282 Human DNA s
35	35	1.2	1257	3	AY070565	AY070565 Drosophil
36	35	1.2	1300	9	BC000794	BC000794 Homo sapi
37	35	1.2	1382	3	AF317731	AF317731 Schistos
38	35	1.2	1452	3	AF380366	AF380366 Schistos
39	35	1.2	1534	9	HSN801279	AL122116 Homo sapi
40	35	1.2	1545	9	AK000609	AK000609 Homo sapi
41	35	1.2	1548	9	BC000784	BC000784 Homo sapi
42	35	1.2	1575	3	AY069691	AY069691 Drosophil
43	35	1.2	1662	9	BC008718	BC008718 Homo sapi
44	35	1.2	1683	3	AF057693	AF057693 Sacculina
45	35	1.2	1764	9	HS4420593	AJ420593 Homo sapi

ALIGNMENTS

RESULT 1
AF184902
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE

AF184902
Mus musculus Osterix (C22) mRNA, complete cds.
AF184902
AF184902.1 GI:18138080
Mus musculus.
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 2960)
Nakashima,K., Zhou,X., Kunkel,G., Zhang,Z., Deng,J.M.,
Behringer,R.R. and de Crombrughe,B.
The novel zinc finger-containing transcription factor osterix is

2960 bp
mRNA
linear
ROD 12-JAN-2002

required for osteoblast differentiation and bone formation
Cell 108 (1), 17-29 (2002)
JOURNAL MEDLINE 21652532
PUBMED 11792318
REFERENCE 2 (bases 1 to 2960)
AUTHORS Nakashima, K., Zhou, X. and de Crombrughe, B.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-1999) Molecular Genetics, M.D. Anderson Cancer
Center, 1515 Holcombe, Houston, TX 77030, USA
FEATURES Location/Qualifiers.
1..2960
source /organism="Mus musculus"
/db_xref="taxon:10090"
/cell_line="C2C12"
/cell_type="myoblasts"
1..2960
/gene="C22"
100..1386
/gene="C22"
/notes="C22"
/note="zinc finger protein; osteoblast/chondrocyte
specific transcription factor"
/codon_start=1
/product="Osterix"
/protein_id="AAL60067.1"
/db_xref="GI:18138081"
/translation="MASSLEEEAHYVSSPLAMLTAAKSKFGSSPLRSTTLKGGT
KPYADLSAPKTMGDAYPAPFSTNGLLSPAGSPAPASGYANDYPPFPSPPTGA
QDGLLVPRGHSSDCLPVSYTLDMTHPYGSIYAGIAGIHPGPNPTWDMHP
GNLGGGGGQDGLQTLSTGPAQPLNPQPTVSDFAPLNPAPYAPHLQPGPO
HVLPODYKPKAVNGSGQLEGGGAAPRRAGTGGSGVAGSAGRSTDCPCOELE
RLGAARAGLRKPIHSHCHIPGGKVKYKASHLKAHLRHTGTPPVVWLFPGCKRFT
SDELRHVTREKFTKLSKRSKRTSHLSKHORTHGEPFGPGPPSPRELGEGR
SVGEENAPRRSSTPAPPEKAHSGSPESNLLEI"
BASE COUNT 647 a 873 c 719 g 721 t
ORIGIN
Query Match 100.0%; Score 2960; DB 10; Length 2960;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2960; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATCTCCCAATTCCTCCCTCCCTCTCCCTCTCCCACTGGCTCGGTTCTCTCC 60
DB 1 ATCTCCCAATTCCTCCCTCCCTCTCCCTCTCCCACTGGCTCGGTTCTCTCC 60
QY 61 ATCTGCTGACTCTCTGGACCGGTCCCGAGCTCGAGGATGGGTCCTCTCTGTTAG 120
DB 61 ATCTGCTGACTCTCTGGACCGGTCCCGAGCTCGAGGATGGGTCCTCTCTGTTAG 120
QY 121 GAAGAAGCTACTATGGCTCCAGTCCCTGGCCATGCTGACTGCAGCTGCAGCAATTT 180
DB 121 GAAGAAGCTACTATGGCTCCAGTCCCTGGCCATGCTGACTGCAGCTGCAGCAATTT 180
QY 181 GCGGCTCTAGCCCTCTGCGGAGCTCAACAACCTTGGGAAAGAGGACACAAAGGCCA 240
DB 181 GCGGCTCTAGCCCTCTGCGGAGCTCAACAACCTTGGGAAAGAGGACACAAAGGCCA 240
QY 241 TACCTGACCTTTCAGCCCAACCAATGAGGAGCCCTACCCAGCTCCCTCTCTCAAGC 300
DB 241 TACCTGACCTTTCAGCCCAACCAATGAGGAGCCCTACCCAGCTCCCTCTCTCAAGC 300
QY 301 ACCAATGACTCTCTCTCTGAGGAGCTCTCCGCGCCAGCCTCTGGCTATGCAAT 360
DB 301 ACCAATGACTCTCTCTCTGAGGAGCTCTCCGCGCCAGCCTCTGGCTATGCAAT 360
QY 361 GACTACCAACCTTTCCTACTCATTTCTGGGCCCCACCGGTGCCCAAGACCTTGGGCTC 420
DB 361 GACTACCAACCTTTCCTACTCATTTCTGGGCCCCACCGGTGCCCAAGACCTTGGGCTC 420
QY 421 CTAGTGCTTAAGGGGACAGCTGCTGACTGCTGCTAGTGTCTACACTTCCCTGGAT 480
DB 421 CTAGTGCTTAAGGGGACAGCTGCTGACTGCTGCTAGTGTCTACACTTCCCTGGAT 480
QY 481 ATGACTCATCCCTATGCTGCTGTGTACAGGAGGATCACCAGGCACTCACCAGGT 540

|||||
481 ATGACTCATCCCTATGGCTGCTGTACAGGAGGATCCACGAGGCATCTCACCAGGT 540
QY 541 CCAGGCAACACACTACTCTTGGTGGACATGACACCTGGGGCAACTGGCTAGTGGT 600
DB 541 CCAGGCAACACACTACTCTTGGTGGACATGACACCTGGGGCAACTGGCTAGTGGT 600
QY 601 GGTACGGGCGAGGTGATGGCTGCAAGGACACTGTGTCACAGGCTTGCACGCTTCCA 660
DB 601 GGTACGGGCGAGGTGATGGCTGCAAGGACACTGTGTCACAGGCTTGCACGCTTCCA 660
QY 661 CTGAACCCCGAGCTGCTTACTTACCATCTGCTTGTCTCCCTTACCCAGCTTCCCTAC 720
DB 661 CTGAACCCCGAGCTGCTTACTTACCATCTGCTTGTCTCCCTTACCCAGCTTCCCTAC 720
QY 721 CCAGGCGCCACCTCTTGAACAGGCGCCAGCATGCTTACCCCAAGATGCTATAAG 780
DB 721 CCAGGCGCCACCTCTTGAACAGGCGCCAGCATGCTTACCCCAAGATGCTATAAG 780
QY 781 CCAGGCGGTGGCAATAGTGGGCAACTGAGGAGGAGTGGTGCAGGCAACCCCTCGG 840
DB 781 CCAGGCGGTGGCAATAGTGGGCAACTGAGGAGGAGTGGTGCAGGCAACCCCTCGG 840
QY 841 GGTGCTGGCACAGGCGGCGGTGATATGCGGCGAGTGGGCGAGGCTTCTACCTGC 900
DB 841 GGTGCTGGCACAGGCGGCGGTGATATGCGGCGAGTGGGCGAGGCTTCTACCTGC 900
QY 901 GACTGCGCCCAACTGTGAGGAGCTAGAGCGGCTGGGGCAGCAGCGGCTGGGCTGAGGAG 960
DB 901 GACTGCGCCCAACTGTGAGGAGCTAGAGCGGCTGGGGCAGCAGCGGCTGGGCTGAGGAG 960
QY 961 AAGCCCATTCACAGCTGCCACATCCCTGGTGGGCAAGGTGTACGGCAAGGCTTCGCAT 1020
DB 961 AAGCCCATTCACAGCTGCCACATCCCTGGTGGGCAAGGTGTACGGCAAGGCTTCGCAT 1020
QY 1021 CTGAAGCCCACTTGGCTGGCACACTGGGAGAGGCTTTCGCTCACTGGCTTTC 1080
DB 1021 CTGAAGCCCACTTGGCTGGCACACTGGGAGAGGCTTTCGCTCACTGGCTTTC 1080
QY 1081 TGGGCAAGAGGTTCACTCGCTCTGAGAGCTGGAGCGCCACAGTGCAGCTCAGCCCGG 1140
DB 1081 TGGGCAAGAGGTTCACTCGCTCTGAGAGCTGGAGCGCCACAGTGCAGCTCAGCCCGG 1140
QY 1141 GAGAAGAAGTTCACTTGGCTGCTGTTCAGAGGCTTTACAGAGCGACCTTTCAGC 1200
DB 1141 GAGAAGAAGTTCACTTGGCTGCTGTTCAGAGGCTTTACAGAGCGACCTTTCAGC 1200
QY 1201 AAACATCAGGCGCACCCAGGAGCGGCGGAGCGCCCGGAGCGCCCGCCAGTGCAGGAG 1260
DB 1201 AAACATCAGGCGCACCCAGGAGCGGCGGAGCGCGCCCGGAGCGCCCGCCAGTGCAGGAG 1260
QY 1261 CTGGGAGGCTGCGAGCTGCGGGAAGAAAGCCCAATCAGCGCCCGGATCTTCCACT 1320
DB 1261 CTGGGAGGCTGCGAGCTGCGGGAAGAAAGCCCAATCAGCGCCCGGATCTTCCACT 1320
QY 1321 TCAGCTGACCCCGGAGGAGGCTTCCAGCTTCCAGGCTTCCAGGCTTCCAGGCTTCCAGG 1380
DB 1321 TCAGCTGACCCCGGAGGAGGCTTCCAGCTTCCAGGCTTCCAGGCTTCCAGGCTTCCAGG 1380
QY 1381 ATCTGAGCGGGTAGAGGAGGCTTCCAGCTTCCAGGCTTCCAGGCTTCCAGGCTTCCAGG 1440
DB 1381 ATCTGAGCGGGTAGAGGAGGCTTCCAGCTTCCAGGCTTCCAGGCTTCCAGGCTTCCAGG 1440
QY 1441 GTGCTGAGACCAATGGTGTGCGGCTCTCTCTATTTGATGCTATCTCTCTCTCTCTCTCT 1500
DB 1441 GTGCTGAGACCAATGGTGTGCGGCTCTCTCTATTTGATGCTATCTCTCTCTCTCTCTCT 1500
QY 1501 TCTCTGTTCCCTTAGGCTATCTCTCTGATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1560
DB 1501 TCTCTGTTCCCTTAGGCTATCTCTCTGATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1560
QY 1561 GTCTTAGCCCAACTCTCTCTCAGGCTTTCAGGCTTTCAGGCTTTCAGGCTTTCAGGCTTTC 1620
DB 1561 GTCTTAGCCCAACTCTCTCTCAGGCTTTCAGGCTTTCAGGCTTTCAGGCTTTCAGGCTTTC 1620

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 37 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 10740: contig of 10740 bp in length
* 10741 10760: gap of unknown length
* 10761 18887: contig of 8127 bp in length
* 18888 18907: gap of unknown length
* 18908 26674: contig of 7767 bp in length
* 26675 26694: gap of unknown length
* 26695 31485: contig of 4791 bp in length
* 31486 31505: gap of unknown length
* 31506 36220: contig of 4715 bp in length
* 36221 36240: gap of unknown length
* 36241 43190: contig of 6950 bp in length
* 43191 43210: gap of unknown length
* 43211 47455: contig of 4245 bp in length
* 47456 52355: gap of unknown length
* 52356 52375: contig of 4880 bp in length
* 52376 57154: contig of 4779 bp in length
* 57155 57174: gap of unknown length
* 57175 63909: contig of 6735 bp in length
* 63910 63929: gap of unknown length
* 63930 70719: contig of 6790 bp in length
* 70720 70739: gap of unknown length
* 70740 73653: contig of 2913 bp in length
* 73653 73673: gap of unknown length
* 73673 79581: contig of 5909 bp in length
* 79582 79602: gap of unknown length
* 79602 83557: contig of 3956 bp in length
* 83558 83577: gap of unknown length
* 83578 89231: contig of 5653 bp in length
* 89231 89250: gap of unknown length
* 89251 93443: contig of 4193 bp in length
* 93444 93463: gap of unknown length
* 93464 97169: contig of 3706 bp in length
* 97170 102251: contig of 5062 bp in length
* 102252 102271: gap of unknown length
* 102272 107091: contig of 4820 bp in length
* 107092 107111: gap of unknown length
* 107112 110480: contig of 3369 bp in length
* 110481 110500: gap of unknown length
* 110501 113418: contig of 2918 bp in length
* 113419 113438: gap of unknown length
* 113439 116688: contig of 3230 bp in length
* 116689 116689: gap of unknown length
* 116690 118403: contig of 1715 bp in length
* 118404 118423: gap of unknown length
* 118424 120544: contig of 2121 bp in length
* 120545 120564: gap of unknown length
* 120565 121993: contig of 1429 bp in length
* 121994 122013: gap of unknown length
* 122014 122930: contig of 917 bp in length
* 122931 122950: gap of unknown length
* 122951 124509: contig of 1559 bp in length
* 124510 124529: gap of unknown length
* 124530 126871: contig of 2322 bp in length
* 126872 126872: gap of unknown length
* 126873 128497: contig of 1625 bp in length
* 128498 128515: gap of unknown length
* 128516 129945: contig of 1433 bp in length
* 129946 129965: gap of unknown length
* 129966 132003: contig of 2034 bp in length
* 132004 132023: gap of unknown length
* 132024 133605: contig of 1582 bp in length
* 133606 133625: gap of unknown length

```

```

* 133626 134100: contig of 475 bp in length
* 134101 134120: gap of unknown length
* 134121 135517: contig of 1397 bp in length
* 135518 135537: gap of unknown length
* 135538 136468: contig of 931 bp in length
* 136469 136488: gap of unknown length
* 136489 137236: contig of 748 bp in length
* 137237 137256: gap of unknown length
* 137257 138860: contig of 1604 bp in length.

FEATURES
      Location/Qualifiers
      source                1..138860
                             /organism="Mus musculus"
                             /strain="C57BL6/J"
                             /db_xref="taxon:10090"
                             /clone="RP23-399N14"
                             /sex="male"
      misc_feature          1..10740
                             /note="assembly_name:Contig270"
      misc_feature          10761..18887
                             /note="assembly_name:Contig269"
      misc_feature          18908..26674
                             /note="assembly_name:Contig268"
      misc_feature          26695..31485
                             /note="assembly_name:Contig267"
      misc_feature          31506..36220
                             /note="assembly_name:Contig266"
      misc_feature          36241..43190
                             /note="assembly_name:Contig265"
      misc_feature          43211..47455
                             /note="assembly_name:Contig264"
      misc_feature          47476..52355
                             /note="assembly_name:Contig263"
      misc_feature          52376..57154
                             /note="assembly_name:Contig262"
      misc_feature          57175..63909
                             /note="assembly_name:Contig261"
      misc_feature          63930..70719
                             /note="assembly_name:Contig260"
      misc_feature          70740..73652
                             /note="assembly_name:Contig259"
      misc_feature          73673..79581
                             /note="assembly_name:Contig258"
      misc_feature          79602..83557
                             /note="assembly_name:Contig257"
      misc_feature          83578..89230
                             /note="assembly_name:Contig256"
      misc_feature          89251..93443
                             /note="assembly_name:Contig255"
      misc_feature          93464..97169
                             /note="assembly_name:Contig254"
      misc_feature          97190..102251
                             /note="assembly_name:Contig253"
      misc_feature          102272..107091
                             /note="assembly_name:Contig252"
      misc_feature          107112..110480
                             /note="assembly_name:Contig251"
      misc_feature          110501..113418
                             /note="assembly_name:Contig250"
      misc_feature          113439..116668
                             /note="assembly_name:Contig249"
      misc_feature          116689..118403
                             /note="assembly_name:Contig248"
      misc_feature          118424..120544
                             /note="assembly_name:Contig247"
      misc_feature          120565..121993
                             /note="assembly_name:Contig246"
                             clone_end:SP6
                             vector_side:left"
      misc_feature          122014..122930
                             /note="assembly_name:Contig245"
      misc_feature          122951..124509
                             /note="assembly_name:Contig244"
      misc_feature          124530..126851

```

Query Match	40.2%	Score 1190;	DB 2;	Length 138860;
Best Local Similarity	99.7%	Pred. No. 0;		
Matches 1644;	Conservative	0;	Mismatches	5;
			Indels	0;
			Gaps	0;
QY 119	AGGAAGAGCTCATATGGCTCCAGTCCCTCGGCATGCTGACTGCAGCCTGCAGCAAT	178		
Db				
QY 25339	AGGAAGAGCTCATATGGCTCCAGTCCCTCGGCATGCTGACTGCAGCCTGCAGCAAT	25180		
Db				
QY 179	TTGGCGGCTCTAGCCCTCTGCGGGACTCAACAACCTCGGGAAGAGGCACAAAGAACG	238		
Db				
QY 25179	TTGGCGGCTCTAGCCCTCTGCGGGACTCAACAACCTCGGGAAGAGGCACAAAGAACG	25120		
Db				
QY 239	CATAGCTGACCTTTCAGCCCCAAAACCATGGGGAGCCCTACCCAGCTCCCTTCTCAA	298		
Db				
QY 25119	CATAGCTGACCTTTCAGCCCCAAAACCATGGGGAGCCCTACCCAGCTCCCTTCTCAA	25060		
Db				
QY 299	GCACCAATGGACTCTCTCTCTCGAGGACGTCTCGGGCCCCAGCCCTCGCATGCAA	358		
Db				
QY 25059	GCACCAATGGACTCTCTCTCTCGAGGACGTCTCGGGCCCCAGCCCTCGCATGCAA	25000		
Db				
QY 359	ATGACTACCCACCCTTCCCTCACTATTCTGCGGCCACCGGTGCCAAGACCCCTGGGC	418		
Db				
QY 24999	ATGACTACCCACCCTTCCCTCACTATTCTGCGGCCACCGGTGCCAAGACCCCTGGGC	24940		
Db				
QY 419	TCCTAGTGCCTAAGGGGCACAGCTCGTCTGACTGCTGCTAGTGTCTACACTTCCCTGG	478		
Db				
QY 24939	TCCTAGTGCCTAAGGGGCACAGCTCGTCTGACTGCTGCTAGTGTCTACACTTCCCTGG	24880		
Db				
QY 479	ATATGACTCATPCCCTATGGCTCGTGGTACAAGGAGGACATCCAGCAGGCATCTCACCAG	538		
Db				
QY 24879	ATATGACTCATPCCCTATGGCTCGTGGTACAAGGAGGACATCCAGCAGGCATCTCACCAG	24820		
Db				
QY 539	GTCAGGCAACACACCTACTCTTGTGGGACATGCACCTGGGGCAACTTGGCTAGGTG	598		
Db				
QY 24819	GTCAGGCAACACACCTACTCTTGTGGGACATGCACCTGGGGCAACTTGGCTAGGTG	24760		
Db				
QY 599	GTGCTCAGGGGCAGGGTGATGGGCTGCAAGGACACATCTGCCAGGGCCCTGCCAGGCCTC	658		
Db				
QY 24759	GTGCTCAGGGGCAGGGTGATGGGCTGCAAGGACACATCTGCCAGGGCCCTGCCAGGCCTC	24700		
Db				
QY 659	CACGTAAACCCCAAGCTGCCTACTTACCACATCTGACTTTGCTCCCTTAAACCCAGCTCCT	718		
Db				
QY 24699	CACGTAAACCCCAAGCTGCCTACTTACCACATCTGACTTTGCTCCCTTAAACCCAGCTCCT	24640		
Db				
QY 719	ACCCAGGCCCCACCTCTTGAACACAGGGCCCCAGCATGCTTACCCCAAGATGTCATATA	778		
Db				
QY 24639	ACCCAGGCCCCACCTCTTGAACACAGGGCCCCAGCATGCTTACCCCAAGATGTCATATA	24580		
Db				
QY 779	AGCCCAAGCGGTTGGCAATAGTGGGCAACTTGGAGGGAGTGGTGCAGCCAAACCCCTC	838		
Db				
QY 24579	AGCCCAAGCGGTTGGCAATAGTGGGCAACTTGGAGGGAGTGGTGCAGCCAAACCCCTC	24520		
Db				
QY 839	GGGTGCTGGCACAGGGGCAGCGGTGATATCGGGCAGTGGGCGAGGGCGTTCTACCT	898		
Db				
QY 24519	GGGTGCTGGCACAGGGGCAGCGGTGATATCGGGCAGTGGGCGAGGGCGTTCTACCT	24460		
Db				
QY 899	CGGACTGCCCCAACTGTCAAGAGCTAGAGCGGCTCGGGGACGACCGGCTGGGCTGAGGA	958		
Db				
QY 24459	CGGACTGCCCCAACTGTCAAGAGCTAGAGCGGCTCGGGGACGACCGGCTGGGCTGAGGA	24400		
Db				
QY 959	AGAAAGCCCATTCACAGTGCACATCCCTGGGTGCGGCAAGGTGTACGCGCAAGGCTTCGC	1018		
Db				
QY 24399	AGAAAGCCCATTCACAGTGCACATCCCTGGGTGCGGCAAGGTGTACGCGCAAGGCTTCGC	24340		
Db				
QY 1019	ATCTAAAGCCCACTTGGCTTGGCACACTGCGGAGAGGCCCTTCGTCTGCAACTGGCTTT	1078		
Db				
QY 24339	ATCTAAAGCCCACTTGGCTTGGCACACTGCGGAGAGGCCCTTCGTCTGCAACTGGCTTT	24280		
Db				

JOURNAL

Submitted (18-APR-2000) Department of Molecular Genetics, Albert Einstein College of Medicine Genome Center, 1300 Morris Park Ave., Bronx, NY 10461, USA
 3 (bases 1 to 138860)
 Grills, G., Han, J., Montgomery, K.T., Lee, E., Long, J., Pomerantz, R., Ioshikhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A., Gordon, M., Goltz, J.S. and Kucherlapati, R.
 Direct Submission
 Submitted (24-APR-2002) Harvard Partners Center for Genetics and Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA 02139, USA
 On Apr 24, 2002 this sequence version replaced gi:18151001.

REFERENCE
 AUTHORS

TITLE
 JOURNAL

COMMENT

-----Genome Center
 Center: Harvard Partners Genome Center
 Center Code: HPGC
 Web site: <http://www.hpcgg.org/Sequence/mouse.html>
 Contact: hpgc@mcg.harvard.edu
 -----Summary Statistics
 Center project name: AAS
 Sequencing vector: pUC18; L08752
 Chemistry: Dye-terminator Big Dye; 100%
 *Consensus quality: 126666 at least Q20
 *Consensus quality: 122952 at least Q30
 *Consensus quality: 116499 at least Q40
 *Estimated insert size: agarose-FP - N/A
 **Estimated insert size: 138140 - sum-of-contigs
 Quality coverage: agarose-FP - N/A
 Quality coverage: 9.1 x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently consists of 37 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 10740: contig of 10740 bp in length
 10741 10760: gap of unknown length
 10761 18887: contig of 8127 bp in length
 18888 18907: gap of unknown length
 18908 26674: contig of 7767 bp in length
 26675 31485: gap of unknown length
 31486 31505: contig of 4791 bp in length
 31506 36220: gap of unknown length
 36221 43190: contig of 4715 bp in length
 43191 43210: gap of unknown length
 43211 47455: contig of 4245 bp in length
 47456 52356: gap of unknown length
 52357 52376: gap of unknown length
 52377 57175: gap of unknown length
 57176 63909: contig of 4779 bp in length
 63910 63929: gap of unknown length
 63930 70719: contig of 6735 bp in length
 70720 73652: gap of unknown length
 73653 73672: gap of unknown length
 73673 79581: contig of 2913 bp in length
 79582 79601: gap of unknown length
 79602 83557: contig of 5909 bp in length
 83558 89230: gap of unknown length
 89231 93463: gap of unknown length
 93464 97169: contig of 5653 bp in length
 97170 102251: gap of unknown length
 102252 102271: gap of unknown length

102272 107091: contig of 4820 bp in length
 107092 107111: gap of unknown length
 107112 110480: contig of 3369 bp in length
 110481 110500: gap of unknown length
 113418 113418: contig of 2918 bp in length
 113419 113438: gap of unknown length
 116668 116688: gap of unknown length
 116689 118403: contig of 1715 bp in length
 118404 118423: gap of unknown length
 118424 120544: contig of 2121 bp in length
 120545 120564: gap of unknown length
 120565 121993: contig of 1429 bp in length
 121994 122013: gap of unknown length
 122014 122930: contig of 917 bp in length
 122931 122950: gap of unknown length
 122951 124509: contig of 1559 bp in length
 124510 124529: gap of unknown length
 124530 126851: contig of 2322 bp in length
 126852 126871: gap of unknown length
 126872 128496: contig of 1625 bp in length
 128497 128516: gap of unknown length
 128517 129949: contig of 1433 bp in length
 129950 129969: gap of unknown length
 129970 132003: contig of 2034 bp in length
 132004 132023: gap of unknown length
 132024 133605: contig of 1582 bp in length
 133606 133625: gap of unknown length
 133626 134100: contig of 475 bp in length
 134101 134120: gap of unknown length
 134121 135517: contig of 1397 bp in length
 135518 135537: gap of unknown length
 135538 136468: contig of 931 bp in length
 136469 136488: gap of unknown length
 136489 137236: contig of 748 bp in length
 137237 137256: gap of unknown length
 137257 138860: contig of 1604 bp in length.

FEATURES
 source

1. 138860
 /organism="Mus musculus"
 /strain="C57BL6/J"
 /db_xref="taxon:10090"
 /clone="RP23-399N14"
 /sex="male"
 1. 10740
 /note="assembly_name:Contig270"
 10761. 18887
 /note="assembly_name:Contig269"
 18908. 26674
 /note="assembly_name:Contig268"
 26695. 31485
 /note="assembly_name:Contig267"
 31506. 36220
 /note="assembly_name:Contig266"
 36241. 43190
 /note="assembly_name:Contig265"
 43211. 47455
 /note="assembly_name:Contig264"
 47476. 52355
 /note="assembly_name:Contig263"
 52376. 57154
 /note="assembly_name:Contig262"
 57175. 63909
 /note="assembly_name:Contig261"
 63930. 70719
 /note="assembly_name:Contig260"
 70740. 73652
 /note="assembly_name:Contig259"
 73673. 79581
 /note="assembly_name:Contig258"
 79602. 83557
 /note="assembly_name:Contig257"
 83578. 89230

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

Consensus quality: 158337 bases at least Q20

Insert size: 184000; agarose-1p

Insert size: 162497; sum-of-contigs

Quality coverage: 3.88 in Q20 bases; agarose-1p

Quality coverage: 4.39 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 43 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1466: contig of 1466 bp in length
* 1467 1566: gap of unknown length
* 1567 2896: contig of 1330 bp in length
* 2897 2996: gap of unknown length
* 2997 4016: contig of 1020 bp in length
* 4017 4116: gap of unknown length
* 4117 5406: contig of 1290 bp in length
* 5407 5506: gap of unknown length
* 5507 6642: contig of 1136 bp in length
* 6643 6742: gap of unknown length
* 6743 8257: contig of 1515 bp in length
* 8258 8357: gap of unknown length
* 8358 10742: contig of 2385 bp in length
* 10743 10842: gap of unknown length
* 10843 12190: contig of 1348 bp in length
* 12191 12290: gap of unknown length
* 12291 13569: contig of 1279 bp in length
* 13570 13669: gap of unknown length
* 13670 15855: contig of 2186 bp in length
* 15856 15955: gap of unknown length
* 15956 17977: contig of 2022 bp in length
* 17978 18077: gap of unknown length
* 18078 20177: contig of 2100 bp in length
* 20178 20277: gap of unknown length
* 20278 21220: contig of 1843 bp in length
* 21221 22220: gap of unknown length
* 22221 24582: contig of 2362 bp in length
* 24583 24682: gap of unknown length
* 24683 27320: contig of 2638 bp in length
* 27321 27420: gap of unknown length
* 27421 29008: contig of 2488 bp in length
* 29009 30008: gap of unknown length
* 30009 32534: contig of 2526 bp in length
* 32535 32634: gap of unknown length
* 32635 35493: contig of 2859 bp in length
* 35494 35593: gap of unknown length
* 35594 37551: contig of 1958 bp in length
* 37552 37651: gap of unknown length
* 37652 41357: contig of 3706 bp in length
* 41358 41457: gap of unknown length
* 41458 43778: contig of 2321 bp in length
* 43779 43878: gap of unknown length
* 43879 47504: contig of 3626 bp in length
* 47505 47604: gap of unknown length
* 47605 50117: contig of 2513 bp in length
* 50118 50217: gap of unknown length
* 50219 53652: contig of 3435 bp in length
* 53653 53752: gap of unknown length
* 53753 56657: contig of 2905 bp in length
* 56658 56757: gap of unknown length
* 56758 60073: contig of 3316 bp in length
* 60074 60174: gap of unknown length
* 60175 64263: contig of 4090 bp in length
* 64264 64363: gap of unknown length
* 64364 68859: contig of 4496 bp in length
* 68860 68959: gap of unknown length
* 68960 72754: contig of 3795 bp in length
* 72755 72854: gap of unknown length
* 72855 77167: contig of 4313 bp in length

* 77168 77267: gap of unknown length
* 77268 81554: contig of 4287 bp in length
* 81555 84716: gap of unknown length
* 84717 84816: contig of 3062 bp in length
* 84817 89925: gap of unknown length
* 89926 90025: contig of 5109 bp in length
* 90026 94331: contig of 4306 bp in length
* 94332 94431: gap of unknown length
* 94432 98564: contig of 4133 bp in length
* 98565 98664: gap of unknown length
* 98665 105959: contig of 7295 bp in length
* 105960 106059: gap of unknown length
* 106060 112344: contig of 6285 bp in length
* 112345 112444: gap of unknown length
* 112445 119806: contig of 7362 bp in length
* 119807 119905: gap of unknown length
* 119906 127259: contig of 7353 bp in length
* 127260 127359: gap of unknown length
* 127360 136297: contig of 8938 bp in length
* 136298 136397: gap of unknown length
* 136398 144884: contig of 8487 bp in length
* 144885 144984: gap of unknown length
* 144985 154492: contig of 9508 bp in length
* 154493 154592: gap of unknown length
* 154593 156697: contig of 12105 bp in length.

FEATURES

source

1..156697
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="UL"
/clone="RP11-147A18"

BASE COUNT 41526 a 40406 c 40099 g 40461 t 4205 others
ORIGIN

Query Match 1.9%; Score 55; DB 2; Length 156697;
Best Local Similarity 100.0%; Pred. No. 4e-20;

Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1941 TGGGATCTGATCCAAAGATGGGTGACCCAGGTCAGGAGCTGCCCCAG 1995

|||||TGGGATCTGATCCAAAGATGGGTGACCCAGGTCAGGAGCTGCCCCAG 150829

RESULT 6

LOCUS

AC073611 192665 bp DNA linear HTG 31-JUL-2002
Homo sapiens chromosome 12 clone RP11-680A11, WORKING DRAFT

DEFINITION

SEQUENCE, 13 unordered pieces.

ACCESSION

AC073611.27 GI:22002370

VERSION

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

KEYWORDS

human.

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 192665)

Muzny, D.M., Adams, C., Ali-Osman, F.R., Allen, C.,

Alsbrooks, S.L., Amarantunga, H.C., Are, J.R., Ayale, M., Banks, T.,

Barbata, J., Benton, J., Blum, K., Blankenburg, K., Bonnin, D.,

Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,

Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,

Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,

Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,

Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,

Devila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,

Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Ding, H.,

Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,

Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,

Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,

Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,

Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,

Harris, C., Harris, K., Hart, M., Haviak, P., Hawes, A., Hernandez, J.,

Hernandez, O., Hodgson, A., Hognes, M., Holloway, C., Hollins, B.,
Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jollivet, S., Joudan, S.,
Karlsson, E., Kelly, S., Khan, U., King, L., Korvan, J., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Lousseged, H.,
Lozad, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mahoney, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M.,
Miner, G., Mear, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G.,
Oganyne, N., Oviedo, R., Pace, A., Payton, B., Peary, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,
Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I.,
Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H., Tang, H.,
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G., and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 192665)
Worley, K.C.

Direct Submission
Submitted (27-JUN-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 192665)
Worley, K.C.

Direct Submission
Submitted (31-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

On Jul 29, 2002 this sequence version replaced gi:21954912.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Drafting Center Code: BCM
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HBMZ
Center clone name: RP11-680A11
----- Summary Statistics
Sequencing vector: Plasmid;
Sequencing vector: M13;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 187403 bases at least Q40
Consensus quality: 190209 bases at least Q30
Consensus quality: 192283 bases at least Q20
Estimated insert size: 176896; sum-of-contigs estimation
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
2183 contig of 2182 bp in length
2183 gap of unknown length
2283 9324: contig of 7042 bp in length
9325 9424: gap of unknown length

* 9425 13352: contig of 3928 bp in length
* 13353 13452: gap of unknown length
* 13453 16937: contig of 3385 bp in length
* 16938 16937: gap of unknown length
* 16938 24070: contig of 7133 bp in length
* 24071 24170: gap of unknown length
* 24171 31951: contig of 7781 bp in length
* 31952 32051: gap of unknown length
* 32052 39982: contig of 7931 bp in length
* 39983 40082: gap of unknown length
* 40083 46308: contig of 6226 bp in length
* 46309 46408: gap of unknown length
* 46409 58146: contig of 11738 bp in length
* 58147 58246: gap of unknown length
* 58247 68637: contig of 10391 bp in length
* 68638 68737: gap of unknown length
* 68738 89001: contig of 20163 bp in length
* 89001 121485: contig of 32485 bp in length
* 121486 121585: gap of unknown length
* 121586 192665: contig of 71080 bp in length.

FEATURES
Location/Qualifiers
1..192665
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"

BASE COUNT 48461 a 47935 c 47370 g 47669 t 1230 others
ORIGIN

Query Match 1.9%; Score 55; DB 2: Length 192665;
Best Local Similarity 100.0%; Pred. No. 4e-20; Indels 0; Gaps 0;
Matches 55; Conservative 0; Mismatches 0;

QY 1941 TGGGATCTGATCCCAAGATGGGTGACCCAGGTCAGGAGGTGCCCCCAG 1995
|||||
DB 168032 TGGGATCTGATCCCAAGATGGGTGACCCAGGTCAGGAGGTGCCCCCAG 167978
|||||

102(5)

RESULT 7
LAU89841
LOCUS
DEFINITION
Lupinus angustifolius diadenosine 5',5'-P1,P4-tetraphosphate
hydrolase mRNA, complete cds.
ACCESSION
U89841
VERSION
U89841.1 GI:1888556
KEYWORDS
SOURCE
Lupinus angustifolius.
ORGANISM
Lupinus angustifolius.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Genisteae;
Lupinus.
REFERENCE
1 (bases 1 to 819)
AUTHORS
Makel, D., Guranowski, A., Ilgoutz, S.C., Moir, A., Blackburn, M.G. and
Gayler, K.R.
TITLE
Cloning and expression of diadenosine 5',5'-P1,P4-tetraphosphate
hydrolase from *Lupinus angustifolius* L
JOURNAL
Biochem. J. 329 (Pt 2), 313-319 (1998)
MEDLINE
98087500
PUBMED
9425114
REFERENCE
2 (bases 1 to 819)
AUTHORS
Gayler, K.R., Maksel, D., Guranowski, A. and Blackburn, G.M.
TITLE
Direct Submission
JOURNAL
Submitted (17-FEB-1997) Biochemistry and Molecular Biology,
University of Melbourne, Parkville, Victoria 3052, Australia
Location/Qualifiers
1..819
/organism="Lupinus angustifolius"
/strain="Unicrop"
/db_xref="taxon:3871"
22..621
/EC_number="3.6.1.17"

CDS

```

BASE COUNT      561 a      244 c      247 g      399 t
ORIGIN
      FNCGTGDEEDWESGESTYL*
Query Match      1.2%; Score 36; DB 3; Length 1451;
Best Local Similarity 100.0%; Pred. No. 5.4e-09;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2925 GATTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2960
      |||||
Db 1402 GATTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1437

```

RESULT	9
BC015812	
LOCUS	
DEFINITION	BC015812 Homo sapiens, Similar to hypothetical protein from clone 24796, 1683 bp mRNA linear PRI 11-OCT-2001
ACCESSION	clone MGC:9533 IMAGE:3920903, mRNA, complete cds.
VERSION	BC015812.1 GI:16041841

SOURCE Homo sapiens.
ORGANISM Homo sapiens.
TITLE Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL 1 (bases 1 to 1683)
AUTHORS Strausberg, R.
REFERENCE Submitted (09-OCT-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT	Contact: MGC help desk Email: cgabs-remail.nih.gov Tissue Procurement: ATCC/DCTD/dmp cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) cDNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada info@bcsc.bc.ca Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fella, Erin Garland, Ran Guin, Leticia Hsiao, Martin Krzywinski, Rita Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saedi, Jacqueline Scheln, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasha van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

```

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 15 Row: m Column: 24.
Location/Qualifiers
1. .1683
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:9533 IMAGE:3920903"
/tissue_type="Skin, melanotic melanoma."
/clone_lib="NIH_MGC_72"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
113. .598
/codon_start=1
/product="Similar to hypothetical protein from clone
24796"
/protein_id="AAH15812.1"
/db_xref="GI:16041842"
/translations="MAKEEPOSIRDLQEOLEKKLSLLIDSFONNSKVVAFMKSPVQGY
LDSHPFLATLLVFTVMSAPVGFLLIVVFTLALLGVILEGLVISVGGSLICI
LCGLGPFVSLAMSGMIIASYVVVSSLLFSPRPLTQQTNSCDFLPAKMSADFEGLYQ
E"
417 a 376 c 400 a 490 t
BASE COUNT

```


Db 2143 GATTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2178

RESULT 13
 AF201905
 LOCUS
 DEFINITION
 Drosophila melanogaster strain Canton-S phosphatase and tensin
 homolog (Pten) mRNA, complete cds.
 ACCESSION
 AF201905
 VERSION
 AF201905.1
 KEYWORDS
 GI:6672083
 SOURCE
 Drosophila melanogaster.
 ORGANISM
 Drosophila melanogaster.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE
 1 (bases 1 to 2255)
 Goberdhan,D.C., Paricio,N., Goodman,E.C., Mlodzik,M. and Wilson,C.
 Drosophila tumor suppressor PTEN controls cell size and number by
 antagonizing the Chico/PI3-kinase signaling pathway
 Genes Dev. 13 (24), 3244-3258 (1999),
 20085000
 PUBMED
 10617573
 REFERENCE
 2 (bases 1 to 2255)
 Goberdhan,D.C.I., Paricio,N., Goodman,E.C., Mlodzik,M. and
 Wilson,C.
 Direct Submission
 Submitted (04-NOV-1999) Biosciences, University of Kent, Giles
 Lane, Canterbury, Kent CT2 7NJ, UK
 JOURNAL
 FEATURES
 Location/Qualifiers
 1. .2255
 /organism="Drosophila melanogaster"
 /strain="Canton-S"
 /db_xref="taxon:7227"
 /chromosomes="2"
 /map="31B-31C"
 /clone="5C"
 1. .2255
 /gene="Pten"
 /note="DPTEN"
 295, 1815
 /gene="Pten"
 /note="lipid/protein phosphatase"
 /codon_start=1
 /product="phosphatase and tensin homolog"
 /protein_id="AAF23236.1"
 /db_xref="GI:6672084"
 /translation="MANTISLSNVTNRNVSKRIRYKKGVDLDITYINDNIITMAY
 PAPDKLEGFLNRNLEDVFLLEENHAQHYKIYNLCSESYDVAKFRGVAVYFPDDH
 PTIELIQRFCSDVDMWLKDSNVVAVHCAGKGTGTCMAYLVFSGIKKASDAL
 AWYDEKTRDKRGVTIPQRRYVQYFSLKVCSSVPYSKVLNVCETRFSESSCVQNL
 MYECSVLVHDSATENAKPDLKLTLPIDFKSEFVLTIKPSIPVSGDVKELPKKSPDK
 IICGHFWLNTFFVRNYSPCSDDGTVNYKIHLKSEIDDVHKDSEHKRFSEEPKISIVF
 EAEFNFSNVOAASEKERNENLVNFSRSDYSLSPNCAEKKVLAIVNDNTTKSQT
 EILKHDTIVTKIQYDSTSKNTSTACKRKQPNKTLPLPSLANDTKEEIKRNHIFNQ
 SIKKTLDIKWQNSEVHITRSINENKNINYSYITCKQSSPKFNCGTEDGEDWESE"
 BASE COUNT 834 a 398 c 400 g 623 t
 ORIGIN
 Query Match 1.2%; Score 36; DB 3; Length 2255;
 Best Local Similarity 100.0%; Pred. No. 5.5e-09;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2525 GATTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2960
 (|||||)|||||
 Db 2206 GATTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2241
 (|||||)|||||
 RESULT 14
 AF022890
 LOCUS
 DEFINITION
 Gallus gallus keratan sulfate proteoglycan mRNA, complete cds.
 ACCESSION
 AF022890

RESULT 12	BC002471	2209 bp	mRNA	linear	PRI 12-JUL-2001
LOCUS	BC002471	2209 bp	mRNA	linear	PRI 12-JUL-2001
DEFINITION	BC002471	2209 bp	mRNA	linear	PRI 12-JUL-2001
ACCESSION	BC002471	2209 bp	mRNA	linear	PRI 12-JUL-2001
VERSION	BC002471.1	2209 bp	mRNA	linear	PRI 12-JUL-2001
KEYWORDS	BC002471.1	2209 bp	mRNA	linear	PRI 12-JUL-2001
SOURCE	BC002471.1	2209 bp	mRNA	linear	PRI 12-JUL-2001
ORGANISM	BC002471.1	2209 bp	mRNA	linear	PRI 12-JUL-2001
REFERENCE	BC002471.1	2209 bp	mRNA	linear	PRI 12-JUL-2001
AUTHORS	BC002471.1	2209 bp	mRNA	linear	PRI 12-JUL-2001
TITLE	BC002471.1	2209 bp	mRNA	linear	PRI 12-JUL-2001
JOURNAL	BC002471.1	2209 bp	mRNA	linear	PRI 12-JUL-2001
REMARK	BC002471.1	2209 bp	mRNA	linear	PRI 12-JUL-2001
COMMENT	BC002471.1	2209 bp	mRNA	linear	PRI 12-JUL-2001
FEATURES	BC002471.1	2209 bp	mRNA	linear	PRI 12-JUL-2001
ORIGIN	BC002471.1	2209 bp	mRNA	linear	PRI 12-JUL-2001

```
VERSION AF022890.1 GI:2570518
KEYWORDS
SOURCE Gallus gallus.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 2455)
AUTHORS Dunlevy, J.R., Neame, P.J., Vergnes, J.-P. and Hassell, J.R.
TITLE Identification of the N-linked oligosaccharide sites in chick
cornel lumanan and keratocan that receive keratan sulfate
J. Biol. Chem. 273 (16), 9615-9621 (1998)
MEDLINE 98211990
PUBMED 9545293
REFERENCE 2 (bases 1 to 2455)
AUTHORS Dunlevy, J.R., Neame, P.J., Vergnes, J.-P. and Hassell, J.R.
TITLE Direct Submission
JOURNAL Submitted (04-SEP-1997) Biochemistry and Molecular Biology,
University of South Florida, 12502, N. Pine Drive, Tampa, FL 33612,
USA
FEATURES
    source      Location/Qualifiers
    1..2455
        /organism="Gallus gallus"
        /db_xref="taxon:9031"
        /dev_stage="1 week old chick"
        139..1200
        /note="keratocan"
        /codon_start=1
        /product="keratan sulfate proteoglycan"
        /protein_id="AAC15506.1"
        /db_xref="GI:2570519"
        /translation="MTLKVCPSSLILLFLVHSVTRVTRQVYNELDPKSHVTFPCP
        QECFSPFNALYCNKGLKEIPAPRIWLYLQNNLIETISEKPFVNATHLRWIN
        LNKNTITNNGIEGVLGSKRLLYLFLEDNELEVPAFLPVGLEQLARNKISRPE
        GVFSNLEFMLDLQNNLDSALQSDTFQGLNSLMQNLAKNSLAKMPLSIPANTLQ
        LFDNNISYFSAIPKVTFLRLNKLSDGIPPPNGFNVSSLDLQLSHNLT
        KIPINAHLEHLHDNRKISVNGTQICPVSIYAEDYGLYGNIPRLRYLRDGNELQ
        PPPLDLMICFQLQAVVI"
BASE COUNT 820 a 462 c 416 g 757 t
ORIGIN

Query Match 1.2%; Score 36; DB 5; Length 2455;
Best Local Similarity 100.0%; Pred. No. 5.5e-09;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2925 GATTTCACAAAAA.....AAAAAAAAAAAAAAAAAAAA 2960
|||||.....
2418 GATTTCACAAAAA.....AAAAAAAAAAAAAAAAAAAA 2453

RESULT 15
AC116967 93491 bp DNA linear HTG 04-APR-2002
LOCUS Dictyostelium discoideum chromosome 2 map 5401525-5495014 strain
DEFINITION AX4, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION AC116967
VERSION AC116967.1 GI:19920066
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyostelid; Dictyostelium.
REFERENCE 1 (bases 1 to 93491)
AUTHORS Gloeckner, G., Eichinger, L., Szafranski, K., Pachebat, J., Dear, P.,
Lehmann, R., Baumgart, C., Parra, G., Parra, J.F., Guigo, R., Kumpf, K.,
Tunggal, B., Cox, E., Quail, M.A., Platzer, M., Rosenthal, A. and
Noegel, A.A.
TITLE Sequence and Analysis of Chromosome 2 of Dictyostelium
JOURNAL Unpublished
REMARK The Dictyostelium Genome Sequencing Consortium
REFERENCE 2 (bases 1 to 93491)
AUTHORS Baumgart, C.
TITLE Direct Submission
JOURNAL Submitted (04-APR-2002) Genome Analysis, Institute of Molecular

COMMENT
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
CDS predictions from Geneid may contain errors. Further Information
is available from IMB Jena, Department of Genome Analysis
(http://genome.imb-jena.de/dictyostelium/)
and the University Cologne, Institute for Biochemistry I
(http://www.uni-koeln.de/dictyostelium/project.shtml
Funding
Agency : Deutsche Forschungsgemeinschaft (DFG).
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
    source      Location/Qualifiers
    1..93491
        /organism="Dictyostelium discoideum"
        /strain="AX4"
        /db_xref="taxon:44689"
        /chromosome="2"
        /map="5401525-5495014"
BASE COUNT 36436 a 9904 c 10019 g 37032 t 100 others
ORIGIN

Query Match 1.2%; Score 36; DB 2; Length 93491;
Best Local Similarity 100.0%; Pred. No. 6.2e-09;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2925 GATTTCACAAAAA.....AAAAAAAAAAAAAAAAAAAA 2960
|||||.....
DB 8360 GATTTCACAAAAA.....AAAAAAAAAAAAAAAAAAAA 8395

Search completed: February 22, 2003, 02:13:15
Job time : 6088 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2003, 14:15:28 ; Search time 5069 Seconds
(without alignments)
16994.339 Million cell updates/sec

Title: US-09-734-329-1
Perfect score: 2960
Sequence: 1 attctccattctccctccc.....aaaaaaaaaaaaaaaaaaaaa 2960

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

arched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg_hum.*
31: em.htg_inv.*
32: em.htg_other.*
33: em.htg_mus.*
34: em.htg_pin.*
35: em.htg_rod.*
36: em.htg_mam.*
37: em.htg_vrt.*
38: em.sy.*
39: em.htgo_hum.*
40: em.htgo_mus.*
41: em.htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match %	Length	DB	ID	Description
1	2960	100.0	2960	10	AF184902	AF184902 Mus muscu
C	2742.6	92.7	138860	2	AC055703	AC055703 Mus muscu
3	1500.8	50.7	2995	9	AF477981	AF477981 Homo sapi
4	1414	47.8	166697	2	AC021103	AC021103 Homo sapi
C	1414	47.8	192865	2	AC073611	AC073611 Homo sapi
6	216	7.3	174437	9	AC018470	AC018470 Homo sapi
7	213	7.2	3471	9	AK056857	AK056857 Homo sapi
C	213	7.2	120169	9	AC005060	AC005060 Homo sapi
C	210.2	7.1	226997	2	AC128798	AC128798 Rattus no
10	207.2	7.0	211696	2	AC099580	AC099580 Mus muscu
C	204	6.9	244434	2	AC126277	AC126277 Mus muscu
12	198.4	6.7	110976	2	AC098301	AC098301 Rattus no
C	180.4	6.1	41507	2	AC014235	AC014235 Drosophil
14	180.4	6.1	159970	3	AC023702	AC023702 Drosophil
C	180.4	6.1	186408	2	AC130390	AC130390 Drosophil
15	180.4	6.1	310364	3	AE003448	AE003448 Drosophil
17	178.8	6.0	3071	3	DME131022	AJ131022 Drosophil
18	174.2	5.9	112334	10	AL606664	AL606664 Mouse DNA
19	171.6	5.8	145528	9	AC003665	AC003665 Homo sapi
20	169	5.7	1254	10	MMU275988	AJ275988 Mus muscu
21	161.2	5.4	1825	10	AF279479	AF279479 Mus muscu
22	161.2	5.4	204996	2	AC097273	AC097273 Mus muscu
23	159.6	5.4	116234	2	AC114075	AC114075 Rattus no
24	159.6	5.4	165594	2	AC118797	AC118797 Rattus no
C	159	5.4	193212	9	AC007405	AC007405 Homo sapi
26	158.8	5.4	591	9	HS330230	AJ330230 Homo sapi
27	154.8	5.2	2071	5	AY057451	AY057451 Fundulus
28	154.6	5.2	2562	5	GA3317960	AJ3317960 Gallus ga
29	150.2	5.1	76376	2	AC021969	AC021969 Homo sapi
30	141.8	4.8	1079	9	AF150628	AF150628 Homo sapi
31	141.8	4.8	1268	9	BC012741	BC012741 Homo sapi
32	141.8	4.8	1351	9	BC010438	BC010438 Homo sapi
33	141.8	4.8	1360	9	BC013946	BC013946 Homo sapi
34	141.8	4.8	1418	9	AF132599	AF132599 Homo sapi
35	140.4	4.7	3062	5	AF388363	AF388363 Danio rer
36	138	4.7	176852	2	AC130082	AC130082 Rattus no
37	137.2	4.6	2637	10	AF283891	AF283891 Mus muscu
38	136.8	4.6	1310	10	MMU275987	AJ275987 Mus muscu
39	136.8	4.6	1378	10	AF251796	AF251796 Mus muscu
40	136.8	4.6	1378	10	AF252285	AF252285 Mus muscu
41	135.4	4.6	1174	10	MMU245644	AJ245644 Mus muscu
42	135	4.6	2721	5	GA3317961	AJ3317961 Gallus ga
43	134.6	4.5	2655	9	HUMTFSP1	J03133 Human trans
44	134.6	4.5	2913	9	AF252284	AF252284 Homo sapi
45	134.4	4.5	1793	5	AY062263	AY062263 xenopus l

ALIGNMENTS

RESULT 1
AF184902
LOCUS AF184902 Mus musculus Osterix (C22) mRNA linear ROD 12-JAN-2002
DEFINITION Mus musculus Osterix (C22) mRNA, complete cds.
ACCESSION AF184902
VERSION AF184902.1 GI:18138080
KEYWORDS
SOURCE Mus musculus.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2960)
REFERENCE Nakashima K., Zhou, X., Kunkel, G., Zhang, Z., Deng, J.M.,
Behringer, R.R. and de Crombrughe, B.
AUTHORS The novel zinc finger-containing transcription factor osterix is
TITLE

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 37 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 10740: contig of 10740 bp in length
* 10741: 10760: gap of unknown length
* 10761: 18887: contig of 8127 bp in length
* 18888: 18907: gap of unknown length
* 18908: 26674: contig of 7767 bp in length
* 26675: 26694: gap of unknown length
* 26695: 31485: contig of 4791 bp in length
* 31486: 31505: gap of unknown length
* 31506: 36220: contig of 4715 bp in length
* 36221: 36240: gap of unknown length
* 36241: 43190: contig of 6950 bp in length
* 43191: 43210: gap of unknown length
* 43211: 47455: contig of 4245 bp in length
* 47456: 52355: contig of 4880 bp in length
* 47476: 52375: gap of unknown length
* 52376: 57154: contig of 4779 bp in length
* 57155: 57174: gap of unknown length
* 57175: 63909: contig of 6735 bp in length
* 63910: 63929: gap of unknown length
* 63930: 70719: contig of 6790 bp in length
* 70720: 70739: gap of unknown length
* 70740: 73652: contig of 2913 bp in length
* 73653: 73672: gap of unknown length
* 73673: 79581: contig of 5909 bp in length
* 79582: 79601: gap of unknown length
* 79602: 83557: contig of 3956 bp in length
* 83558: 83577: gap of unknown length
* 83578: 89230: contig of 5653 bp in length
* 89231: 89250: gap of unknown length
* 89251: 93443: contig of 4193 bp in length
* 93444: 93463: gap of unknown length
* 93464: 97169: contig of 3706 bp in length
* 97170: 97189: gap of unknown length
* 102251: 102251: contig of 5062 bp in length
* 102252: 102271: gap of unknown length
* 102272: 107091: contig of 4820 bp in length
* 107092: 107111: gap of unknown length
* 107112: 110480: contig of 3369 bp in length
* 110481: 110500: gap of unknown length
* 110501: 113418: contig of 2918 bp in length
* 113419: 113438: gap of unknown length
* 113439: 116668: contig of 3230 bp in length
* 116669: 116688: gap of unknown length
* 116689: 118403: contig of 1715 bp in length
* 118404: 118423: gap of unknown length
* 118424: 120544: contig of 2121 bp in length
* 120545: 120564: gap of unknown length
* 120565: 121993: contig of 1429 bp in length
* 121994: 122013: gap of unknown length
* 122014: 122930: contig of 917 bp in length
* 122931: 122950: gap of unknown length
* 122951: 124509: contig of 1559 bp in length
* 124510: 124529: gap of unknown length
* 124530: 126851: contig of 2322 bp in length
* 126852: 126871: gap of unknown length
* 126872: 128496: contig of 1625 bp in length
* 128497: 128516: gap of unknown length
* 128517: 129949: contig of 1433 bp in length
* 129950: 129969: gap of unknown length
* 129970: 132003: contig of 2034 bp in length
* 132004: 132023: gap of unknown length
* 132024: 133605: contig of 1582 bp in length
* 133606: 133625: gap of unknown length

```

```

* 133626 134100: contig of 475 bp in length
* 134101 134120: gap of unknown length
* 134121 135517: contig of 1397 bp in length
* 135518 135537: gap of unknown length
* 135538 136468: contig of 931 bp in length
* 136469 136488: gap of unknown length
* 136489 137236: contig of 748 bp in length
* 137237 137256: gap of unknown length
* 137257 138860: contig of 1604 bp in length.

FEATURES
      Location/Qualifiers
      1..138860
        /organism="Mus musculus"
        /strain="C57BL6/J"
        /db_xref="taxon:10090"
        /clone="RP23-399N14"
        /sex="male"
      1..10740
        /note="assembly_name:Contig270"
      misc_feature
      10761..18887
        /note="assembly_name:Contig269"
      misc_feature
      18908..26674
        /note="assembly_name:Contig268"
      misc_feature
      26695..31485
        /note="assembly_name:Contig267"
      misc_feature
      31506..36220
        /note="assembly_name:Contig266"
      misc_feature
      36241..43190
        /note="assembly_name:Contig265"
      misc_feature
      43211..47455
        /note="assembly_name:Contig264"
      misc_feature
      47476..52355
        /note="assembly_name:Contig263"
      misc_feature
      52376..57154
        /note="assembly_name:Contig262"
      misc_feature
      57175..63909
        /note="assembly_name:Contig261"
      misc_feature
      63930..70719
        /note="assembly_name:Contig260"
      misc_feature
      70740..73652
        /note="assembly_name:Contig259"
      misc_feature
      73673..79581
        /note="assembly_name:Contig258"
      misc_feature
      79602..83557
        /note="assembly_name:Contig257"
      misc_feature
      83578..89230
        /note="assembly_name:Contig256"
      misc_feature
      89251..93443
        /note="assembly_name:Contig255"
      misc_feature
      93464..97169
        /note="assembly_name:Contig254"
      misc_feature
      97190..102251
        /note="assembly_name:Contig253"
      misc_feature
      102272..107091
        /note="assembly_name:Contig252"
      misc_feature
      107112..110480
        /note="assembly_name:Contig251"
      misc_feature
      110501..113418
        /note="assembly_name:Contig250"
      misc_feature
      113439..116668
        /note="assembly_name:Contig249"
      misc_feature
      116689..118403
        /note="assembly_name:Contig248"
      misc_feature
      118424..120544
        /note="assembly_name:Contig247"
      misc_feature
      120565..121993
        /note="assembly_name:Contig246"
      clone_end.SP6
      vector_side:left"
      122014..122930
        /note="assembly_name:Contig245"
      misc_feature
      122951..124509
        /note="assembly_name:Contig244"
      misc_feature
      124530..126851

```

```
/note="assembly_name:Contig243"
128872..128496
/note="assembly_name:Contig242"
128517..129949
/note="assembly_name:Contig241"

Query Match      92.78; Score 2742.6; DB 2; Length 138860;
Best Local Similarity 97.88; Pred. No. 0;
Matches 2815; Conservative 0; Mismatches 40; Indels 22; Gaps 2;

QY 63 CTGCTGACTCTCTGGGACCGGTCGCCAGCTCGAGATGGCGCTCTCTGCTTGAGGA 122
  || || || || || || || || || || || || || || || || || || || || ||
Db 25295 CTCCTCTCCCTAAGCAGTGCCTCTCTCCCGCCCGCCCTCTCTCTACAGGA 25236

QY 123 AGAAGCTACATGCTCCATGCTCCCTGGCCATGCTGACTGACGCTGCACGAATTTGG 182
  || || || || || || || || || || || || || || || || || || || || ||
Db 25235 AGAAGCTACATGCTCCATGCTCCCTGGCCATGCTGACTGACGCTGCACGAATTTGG 25176

  183 CGGCTCTAGCCCTCTGCGGGACTCAACAACCTCGGGAAGGAGGCAAAAGAGCCATA 242
  || || || || || || || || || || || || || || || || || || || || ||
Db 25175 CGGCTCTAGCCCTCTGCGGGACTCAACAACCTCGGGAAGGAGGCAAAAGAGCCATA 25116

QY 243 CGCTGACCTTTACGCCCCCAAAACATAGGGGAGCGCTTACCAGCTCCCTTCTCAAGCAC 302
  || || || || || || || || || || || || || || || || || || || || ||
Db 25115 CGCTGACCTTTACGCCCCCAAAACATAGGGGAGCGCTTACCAGCTCCCTTCTCAAGCAC 25056

QY 303 CAATGGACTCTCTCTCTGAGGAGTCTCTCGGCGCCCGCCAGCGCTTGGCTATGCAAAATGA 362
  || || || || || || || || || || || || || || || || || || || || ||
Db 25055 CAATGGACTCTCTCTCTGAGGAGTCTCTCGGCGCCCGCCAGCGCTTGGCTATGCAAAATGA 24996

QY 363 CTACCCACCCCTCCCTCACTATTTCTTGGGCGCCACCGGTGCCCAAGACCTTGGGCTCT 422
  || || || || || || || || || || || || || || || || || || || || ||
Db 24995 CTACCCACCCCTCCCTCACTATTTCTTGGGCGCCACCGGTGCCCAAGACCTTGGGCTCT 24936

QY 423 AGTGCTAAGGGGCACAGCTGCTGACTGCTGCTAGTGTCTACACTTCCCTGGATAT 482
  || || || || || || || || || || || || || || || || || || || || ||
Db 24935 AGTGCTAAGGGGCACAGCTGCTGACTGCTGCTAGTGTCTACACTTCCCTGGATAT 24876

QY 483 GACTCATCCCTATGGCTGGTGTACAGGAGGATTCACAGCGCTTCCAGAGTCC 542
  || || || || || || || || || || || || || || || || || || || || ||
Db 24875 GACTCATCCCTATGGCTGGTGTACAGGAGGATTCACAGCGCTTCCAGAGTCC 24816

QY 543 AGGCAACACACTACTCTCTGTTGGGACATGCACCTCGGGGGCAACTGGCTAGTGGTGG 602
  || || || || || || || || || || || || || || || || || || || || ||
Db 24815 AGGCAACACACTACTCTCTGTTGGGACATGCACCTCGGGGGCAACTGGCTAGTGGTGG 24756

  603 TCAGGGCCAGGGTGATGGGCTGCAAGGACACTGTCCACAGGCGCTGCCAGCTCCACT 662
  || || || || || || || || || || || || || || || || || || || || ||
Db 24755 TCAGGGCCAGGGTGATGGGCTGCAAGGACACTGTCCACAGGCGCTGCCAGCTCCACT 24696

QY 663 GAACCCCGAGTGGCTACTTACCCTATGCTTGGCTCCCTTTAACCCAGCTCCCTACCC 722
  || || || || || || || || || || || || || || || || || || || || ||
Db 24695 GAACCCCGAGTGGCTACTTACCCTATGCTTGGCTCCCTTTAACCCAGCTCCCTACCC 24636

QY 723 AGCGCCCGACCTCTTGCAACAGGCGCCCGCAGCATGCTCTACCCCAAGATGCTATAGCC 782
  || || || || || || || || || || || || || || || || || || || || ||
Db 24635 AGCGCCCGACCTCTTGCAACAGGCGCCCGCAGCATGCTCTACCCCAAGATGCTATAGCC 24576

QY 783 CAAGCGGTTGGCAATAGTGGGCAACTGGAGGGAGTGGTGCAGCCAAACCCCTCGGG 842
  || || || || || || || || || || || || || || || || || || || || ||
Db 24575 CAAGCGGTTGGCAATAGTGGGCAACTGGAGGGAGTGGTGCAGCCAAACCCCTCGGG 24516

QY 843 TGCTGGCACAGGGGACCGGTGATATGCGGGCAGTGGGCGAGGGGTTTACTGCGGA 902
  || || || || || || || || || || || || || || || || || || || || ||
Db 24515 TGCTGGCACAGGGGACCGGTGATATGCGGGCAGTGGGCGAGGGGTTTACTGCGGA 24456

QY 903 CTGCCCCAACGTGTCAGGAGCTAGAGCGCTCGGGGACGACGCGCTGGGCTGAGGAAGAA 962
  || || || || || || || || || || || || || || || || || || || || ||
Db 24455 CTGCCCCAACGTGTCAGGAGCTAGAGCGCTCGGGGACGACGCGCTGGGCTGAGGAAGAA 24396

QY 963 GCCCATTCACAGCTGCCACATCCCTGGGTGGGCAAGGTGTACGCAAGGCTTCGCATCT 1022
  || || || || || || || || || || || || || || || || || || || || ||
Db 24395 GCCCATTCACAGCTGCCACATCCCTGGGTGGGCAAGGTGTACGCAAGGCTTCGCATCT 24336
```


[illegible]

Qy	1779	CTTCCGTGAGCTTTGGTTTTCTTTTTTAAACAACACGATGATGATGAATGATGATG	1838
Db	150628	ACACCATTTGCTTCATCTCCTCAAGAACTCTTTTCTTTTACAAACAATAATG	150687
Qy	1839	ATAATTTATTGCCCTCGTGTTCTTCATTAGGAACACAGATTAGGAGATTGGTGTAG	1898
Db	150888	ATAATTTATTGCCCTCGTGCGCT-----CTTCATCAGGGTATTGGGGTAG	150736
Qy	1999	TAACTGGCCGGGAGCAGAGTGCCAAGAAGGGGAAGTCCAATGGGGATCTGATCCCAA	1958
Db	150737	TGACCTGGCCAG---AGGGTGCCAAGAGGGGGCAGACCAGTGGGATCTCATCCCAA	150792
Qy	1959	GATGGGTGACCCAGGCTCAGGGAGGCTGCCGCCAGCCTTGAGTACTTAACCCCTATGC	2018
Db	150793	GATGGGTGACCCAGGCTCAGGGAGGCTGCCGCCAGCCTGTATATTTAACCCCTATGT	150852
Qy	2019	GCAGGAGTAGAAGAATAGTAGTAATAATAATAATCTATTATCTAAGTATGCA	2078
Db	150853	ACAGGAGTAGTAAGTAG-----TAATAATCTATTATGTAAAGTATGCA	150899
Qy	2079	TGACGGGTCAAGTACAGTGAAGTGGAGGGGAAGGATCTCCCGCCCCCAAGGAAT	2138
-	150898	TGACGGGTCAAGTACAGTGAAGTGGGGAGGAGTGSAT-CCAATTCGTCTAAGRAAT	150955
-	2139	TCTAGTCAATGATCTCTGTATAGAC-AAATGATAGTGGAGACCTTGCTCGTA---GA	2193
Db	150956	TCTAGTCAATGATCTCTGTATAGACAAAATGTTAGTGGAGAAGATCTTGTTAATAGAA	151015
Qy	2194	TTCTATCCFCGAGGCTCG-----AGAGTTCTTTTCAGTTGAGTTTGGGTGTTTC	2248
Db	151016	TGCTATCATCAGAACTCAGTTGATGAGGTTTCTCTGTAATGAAGTCTCACAAATG	151075
Qy	2249	GG-----CCTCTTTTAGTATTCTGTGGGTGTCTCTGTTAGCGAGTCACTAA	2297
Db	151076	GTTTAGTACATCTCTGCTAAACAGTTGATGGGTATCTCTGATTAGGGGATCCCTAA	151135
Qy	2298	GATCCCAGCCCGACGACGAAAGCTGTGAACCTTCAAGTCTTATGCGGGG---AGGACT	2354
Db	151136	TATCCCAGCCCGACGCGAG-AAGCTGTGAACCTCAAGTCTTATGAGGGGAGAAGACT	151194
Qy	2355	GGAATGPACCCAGTCTCTCGACCCGACTGCAGATCAGTTGCCCTCCCTGATCCTCTC	2414
Db	151195	GGAATGPACCCCATCTCCCT----TGACTGCAGACAGGTTCCCTCAGTCCCCACCCC	151249
Qy	2415	TCATACCCCTGACCTCACAGGTTATCCCTTGTCTCATGTTACAGACAGCTTGCAG	2474
Db	151250	TTAGACACCATGACCCCATCAGTTAATCCCTGTTGCCATGGTTATGAGAGCTTGCAG	151309
Qy	2475	CTGCCATCTTAACGTGCTCTTTGGGGGAGAGCCCACTTAACAGGAGGATTTGTTGG	2534
-	151310	CTGCCATCTTAGATGTGCTTTTGGG--AAGCCCATCTAACAGGAGGACATTTGTTGG	151367
-	2535	AGTGCCCTCCTGAAAAAGTAGTGGCGGCAAGGCTTCTCTGGGATCAAAATCAATAA	2594
Db	151368	GGGTGACCTCTCTGAAGAA--TGGGTGGGGAAGGCTTCTCTAGGATCAGATTCAAAT-	151423
Qy	2595	ATCAAGTATTATTGAATGCTTAATGTGCAAGGC-----CTGGTGCTCTAGA	2642
Db	151424	---AAGTATGATTAGTGCCCTACTCTGTGCAAGGCNCTATGCTAGATCTGGTGCTAGA	151480
Qy	2643	AGCCACGAGAAAGAAATTA-----TAACAGGACAGAGTCCCTAAACTAA-----	2687
Db	151481	AGCCCTGAGAAAGAACTTAAGAGCTAGGAGGACAGAGGCCCCCAAGCTGATCTGGTGT	151540
Qy	2688	ACATCCACAGCCCCCAATCTAGGAGGTTCA-----CTCCATCTCAGTGACTTT	2737
Db	151541	GCAFTCCACGACCCCCACCTCGGGACTTTGGATGCTCCCATCTCCACCTCCCACTGAC	151600
Qy	2738	TAAAGCGGCTTGTGCTCTTGAATGCCCTTCTCTGAGATTTTGGATCTTCCCTGTTCTGT	2797
Db	151601	TAAAGCGGCTCTGCGCTT-----CCTGTAAAGTTGATCTCCTCTTCTTCTGT	151648
Qy	2798	CCCTGCTCCTCTTAGCGCTCAAGATAAAGGGTAAA---GCCATGGAGTCTGGGAAG	2853

```

|||||  || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 151649 CCCCCTGCTGTCCTCAAGGCCCCCAAGTTAAAGGGTTTAAAGCGCTTGGAAGAGAA 151708

Qy 2854 CATACATCCGTTGACGGGATCGTCCCTTGTGCGAATCTTCTCTTTTTTAAATTAATA 2913
||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 151709 CATGTGGAATGAAGGGATCATGCCCTTGTGTGA---GTCTTTTTTAAATTAATA 151765

Qy 2914 AATAAAAAGTTCGATTTTCAA 2933
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 151766 AATAAAAAGTTGGATTTCGAA 151785

RESULT 5
AC073611/c
LOCUS AC073611 192665 bp DNA linear HTG 31-JUL-200
DEFINITION Homo sapiens chromosome 12 clone RP11-680A11, WORKING DRAFT
SEQUENCE, 13 unordered pieces.
AC073611
AC073611.27 GI:22002370
HTG; HTGS PHASE1; HTGS DRAFT; HTGS FULLTOP.
```

TITLE
JOURNAL
REFERENCES
AUTHORS
TITLE
JOURNAL


```
QY 1153 ACTTGCCCTGCTCTGTTCCAGCGCTTTACAGAGAGCGACCTTTGACACACATCAGCGC 1212
D 168826 ACCTGCTGCTCTGCTCTCAAGCGCTTTACCGAAGCGACCACTTGAGCAACACAGCGC 168767
QY 1213 ACCACGGGGAGCGAGCGCGGACCGCCCGCCCAAGTGGCCCTTAAGAGAGCTGGGGAGGGT 1272
D 168766 ACCCATGGAAACAGCGCGGGTCCCTCCCAAGTGGCCCAAGAGAGCTGGGGAGGGC 168707
QY 1273 CGCAGCTCGGGGAGAGAACCAATCAGCGCGCCCGATCTCCCACTTGCCTGCGACCC 1332
D 168706 CGCAGCAAGGGGGAGAGAGCGAGTCAAGCGCGCCCGACCTTCTGCTCGCCAGCAACC 168647
QY 1333 CCAGAAAAGCCCGACGAGGCGAGCGCCAGAGCAGAGCAACCTGCTAGAGATCTGAGCGGG 1392
D 168646 CCAGAAAAGCCCGTGGAGGAGCGCTGAGCAGAGCAACTTCTGGAGATCTGAGCGGG 168587
QY 1393 TAGAGGAAGTCTCCAGCTCCAGGCTCTCTTCCAGGCTCTCTTGGCGTCTGGAACCA 1452
D 168586 T---GGAAGTCTCCACCGCGGCTGCCCTGACAGTCTCTTGGCTCTCTAGA---CCA 168531
QY 1453 TTGCTGCGCCCTCGCTCTCT---CCTATTGATGCTATCTCTGGG----- 1496
D 168530 CTGCTTGCCAACTACTCTCTTACCCCATGCATGCCATCTCTTCGGGGCTCTCTCCCTCTG 168471
QY 1497 -----GCTCTCTCTGTTCCCTAGGCTATCTCTTGCATGCTCTCTCAGTT---CTTC 1546
D 168470 TCTCCCTCTGGCCATCTCTGGGCTTGGGATCTCTCTTGCATGCTCTCTCAGCTCACCTTC 168411
QY 1547 TCTCTTTGTCAGAGTCT-----TAGCCAACTCTCTCAGGCTTTTGGC-AGTGCCT 1598
D 168410 TCTCTTACCATGAGCTGGCTTTTCCACAACTCTCTATCTCAGGCGCTCCCTTGTGCT 168351
QY 1599 AGTTCCTATGCTCGAGCTCTCAACTTTTCTCTCTGCGCCCTGCTTCTCACAGCTTCC 1658
D 168350 GATACCTGCTCACTCGGGTCTCTAGACTCTGGCCCTGCGCACCAACACACTTTCTATTG 168291
QY 1659 ATCTGGGCTCATCATTTCTCTAATTAAGTCTGTTGCCATCTAATCTTCTGCTTCCCAAT 1718
D 168290 GGCTCCCAACACTATTCTCCATCTCACTCTTGAGATGATACCCCTTCTGCTCTCAAG 168231
QY 1719 CTAATTGCGCTTTCCGAGCTTCCAGGCTGTCGCTCGATTCGCCCGCCACCTTTCGT 1778
D 168230 CTTATTCTGCTCTGCTCAGCTCAGCTCAGGCTTCCAGTCTT-----CCCACTTCTT 168180
QY 1779 CTTCCTGAGCTTCTGTTCTTTCTTTTAAACAACAGATGATGATGATGATG 1838
D 168179 ACACCATGCTTTCCATCTCCAGAACTCTTTTCTCTTTTACAAAACAAATGATAATG 168120
QY 1839 ATAATTTATTTGCCCTGGTGTCTTCTATTAGGAACACAGAGTTAAGGAGATTTGGTGTAG 1898
D 168119 ATAATTTATTTGCCCTGGTGTCTT-----CTTCACTAGGGGTATTGGGGTATG 168071
QY 1899 TAACCTGCGGGAGCAGAGTGCCAGAGGGGGAAGTCCCAATGGGATCTGATCCAAA 1958
D 168070 TGACCTGCGCCAG---AGGGTGCCAGAGGGGGGCGACCACTGGGGATCTGATCCAAA 168015
QY 1959 GATGGGTGACCCAGGGTCAAGGAGGCTGCCCGCCAGCTTGTAGTACTTAACCCCTATGC 2018
D 168014 GATGGGTGACCCAGGGTCAAGGAGGCTGCCCGCCAGCTTGTATATTAAACCCCTATGT 167955
QY 2019 GCCAGGAGTAAGAAATAGTAATAATAATAATAATTAATTTATTTATCTAAGTTATGA 2078
D 167954 ACCAGGAGTATGATAG---TAATTAATCTATTATTTATGATGATGATGA 167910
QY 2079 TGACGGGTGAGTACAGTACAGTGGAGGGAAGGAGTCTCCCGCCCGCCCAAGGAAT 2138
D 167909 TGACGGGTGAGTACAGTACAGTGGAGGGAAGTGGAT---CCATTTCTGCTAAGGAAT 167852
QY 2139 TCTAGTCAATGCACTCTCTGATAGAC-AAATGATAGTGGAGACCTTGTCTCGTA----GA 2193
D 167851 TCTAGTCAATGCACTCTCTGATAGCAAAATGTTAGTGGAGAGATCTTGTTAATAGAA 167792
```

```
QY 2194 TTCTATCTCGAGGCTCTCCG-----AGAGTTTCTTTTTCAGTTTCAGTTTGGGTTGTTCT 2248
D 167791 TGCTATCATCAGAACTCAGTTTGTATAGGGTTTCTCTTGTATGAAGTCTCTACAAATTTG 167732
QY 2249 GG-----CCTCTTTTAGAGTTTCTGTGGTGTCTCTCTGTTAGGAGTCACTAA 2297
D 167731 GGTATGCTACATCTCTCTCTAAACAGTTGATGGGTATCTCTTGTATAGGAGTCCCTAA 167672
QY 2298 GATCCCGAGCGCCAGCAGAAAGCTGTGAACCTTCAAGTCTCTATGGGGGG---AGGACT 2354
D 167671 TATCCCGAGCGCCAGCGAG---AAGCTGTGAACCTTCAAGTCTCTATGGAGGAGAGGACT 167613
QY 2355 GGAATGTACCCAGTCTCTCGACCCGAGTGCAGATCAGGTTCTCCCTCTCTCTCTTC 2414
D 167612 GGAATGTACCCAGTCTCTCT-----TGACTGCAGAGAGGTTCTCTCACTGCCCAACCC 167558
QY 2415 TCATACCTCTGACCTCACCAGGTTATCCCTCTCTCTCTGCTCATGTTTACAGAGAGCTTGCG 2474
D 167557 TTAGACACCATGACCCCATCAGGTTTATCCCTCTTGCATGTTTATGGAGAGCTTGCGAG 167498
QY 2475 CTGCGATCTTAAAGTCTCTTTGGGGGAGAGCCCACTTAACAGAGAGATTTTGGTTGG 2534
D 167497 CTGCGATCTTAAAGTCTCTTTGGGG--AAGCCCATCTTAACAGAGAGATTTGTTGG 167440
QY 2535 AGGTGCGCTCTCTGAAGAGTGTGGGCAAGGCTTCTCTGGGATCAAAATTTCAAAATA 2594
D 167439 GGGTGCACTCTCTGAAGAA--TGGGTGGGAAGGCTTCTCTAGGATCAGATTTCAAAAT-- 167384
QY 2595 ATCAAGTATTATTGAATGCTTAATATGTGCAAGGC-----CTGGTGCTCTAGA 2642
D 167383 ---AAGTATGTTAGTGCTCTCTGTGCAAGGACTATGCTAGATCTGCTGGTCTAGA 167327
QY 2643 AGCCACGAGAAAGATTTA-----TAACAGGACAGAGTCCCTAAACTAA- 2687
D 167326 AGCCCTGAGAAGAACTTAAAGAGCTAGAGGACAGAGGCGCCCAAGCTGATCTGGTGT 167267
QY 2688 ACATCCACAGCGCCCAATCTAGAGGCTTCA-----CTCCATCCAGTACTT 2737
D 167266 GCATCCAGGACCGCCCACTGGAGCTTGGATGCTCCCATCTCCACTCCAGTACTT 167207
QY 2738 TAAAGCGCTTGTGCTTTGAAATGCTTCTCTGAGATTTTGGATCTTCTCTGTTCTGT 2797
D 167206 TAAAGCGCTTGTGCTTT-----CCTGTAAGTGTGATCTCTCTTCTGT 167159
QY 2798 CCGCTGCTCTCTAGGCTCAAGATAAAGGTTAA-----GCCATGAGTCTGGAGAGAG 2853
D 167158 CCGCTGCTCTCTAAGGCCCAAGTTAAAGGTTTAAAGCGCTGGAGCTGGGAGAGAA 167099
QY 2854 CATAACGCTCTGAGCGGATCGTCCCTTGTGGAATCTTCTTTTAAATTTAATA 2913
D 167098 CATTGGAATGGAAGGATCATGCTTGTGGA---GCTTTTTTTTAAATTTAATA 167042
QY 2914 AATAAAGTTGATTTTCAAA 2933
D 167041 AATAAAGTTGATTTGAAA 167022
```

RESULT 6
AC018470
LOCUS Homo sapiens BAC clone Rpl1-572N21 from 2, complete sequence.
DEFINITION AC018470
ACCESSION AC018470
VERSION AC018470.11 GI:15144488
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 174437)
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792

PUBMED REFERENCE	2. .131	repeat_region	2. .131	repeat_region
2 (bases 1 to 174437)	/rpt_family="L1"	misc_feature	4. .22	misc_feature
Kang, K. and Haakenson, W.	/note="similar to Homo sapiens EST BE891364 (NID:g10350624)"	repeat_region	129. .161	repeat_region
The sequence of Homo sapiens BAC clone RP11-572N21 unpublished	/rpt_family="(T)n"	repeat_region	132. .442	repeat_region
3 (bases 1 to 174437)	/rpt_family="Alu"	misc_feature	135. .162	misc_feature
Waterston, R.H.	/note="similar to Homo sapiens EST BE891364 (NID:g10350624)"	misc_feature	138. .152	misc_feature
Direct Submission	/note="similar to Homo sapiens EST BE960930 (NID:g10571635)"	repeat_region	140. .156	repeat_region
Submitted (09-AUG-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA	/note="similar to Homo sapiens EST AA634090 (NID:g2557304) ac74ell.s1"	repeat_region	443. .486	repeat_region
4 (bases 1 to 174437)	/rpt_family="L1"	repeat_region	487. .775	repeat_region
Waterston, R.	/rpt_family="Alu"	misc_feature	619. .635	misc_feature
Direct Submission	/note="similar to Homo sapiens EST BG490821 (NID:g13452331)"	repeat_region	776. .1642	repeat_region
Submitted (07-NOV-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA	/rpt_family="L1"	misc_feature	778. .794	misc_feature
On Aug 9, 2001 this sequence version replaced gi:13399438.	/note="similar to Homo sapiens EST BE891364 (NID:g10350624)"	misc_feature	779. .794	misc_feature
----- Genomex Center	/note="similar to Homo sapiens EST AL517609 (NID:g12781102)"	misc_feature	782. .794	misc_feature
Center: Washington University Genome Sequencing Center	/note="similar to Homo sapiens EST BE960930 (NID:g10571635)"	repeat_region	1655. .1681	repeat_region
Center code: WUGSC	/rpt_family="(T)n"	misc_feature	1655. .1670	misc_feature
Web site: http://genome.wustl.edu/gsc	/note="match to EST AW449728 (NID:g9990504)"	misc_feature	1657. .1681	misc_feature
Contact: saplens@wustl.wustl.edu	/note="similar to Homo sapiens EST BE891364 (NID:g10350624)"	repeat_region	1966. .2007	repeat_region
----- Summary Statistics	/rpt_family="AT-rich"	repeat_region	1968. .2138	repeat_region
-----	/rpt_family="L1"	repeat_region	2189. .2315	repeat_region
Center project name: H_NH0572N21	/rpt_family="Alu"	repeat_region	2316. .2486	repeat_region
-----	/rpt_family="L1"	repeat_region	2487. .2734	repeat_region
NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.	/rpt_family="Alu"	repeat_region	2785. .3280	repeat_region
This sequence was finished as follows unless otherwise noted:	/rpt_family="L1"	repeat_region	2796. .2827	repeat_region
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.	/rpt_family="AT-rich"	repeat_region	3281. .3495	repeat_region
MAPPING INFORMATION:	/rpt_family="MERL-type"	misc_feature	3409. .3411	misc_feature
The RP11-11 human BAC library was made from the blood of one male donor, as described by Oseogawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenno, M., Catanesi, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)	/note="similar to Homo sapiens EST BF698554 (NID:g11983962)"	repeat_region	3496. .3556	repeat_region
VECTOR: pBAC3.6	/rpt_family="L1"	misc_feature	3532. .3565	misc_feature
NEIGHBORING SEQUENCE INFORMATION:	/note="similar to Homo sapiens EST AL514989 (NID:g12778482)"	misc_feature	3554. .3574	misc_feature
The clone sequenced to the left is RP11-451F14, 200 bp overlap; the clone sequenced to the right is RP11-483E17. Actual start of this clone is at base position 1 of RP11-572N21; actual end is at base position 174437 of RP11-572N21.	/note="similar to Homo sapiens EST AA648682 (NID:g2575111) ns1ld09.s1"	repeat_region	3557. .3825	repeat_region
Location/Qualifiers	/rpt_family="Alu"	misc_feature	3557. .3568	misc_feature
1. 174437				
/organism="Homo sapiens"				
/db_xref="taxon:9606"				
/chromosome="2"				
/map="2"				
/clone="RP11-572N21"				
/clone_lib="RPC1-11"				

Homsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvan, J., Kovar, C.,
 Kravovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H.,
 Lozado, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mathiney, E., McLeod, M. P., Meador, M., Mei, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G.,
 Oraguyne, N., Oviado, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,
 Scherer, S., Scott, G., Shen, H., Shooshari, N., Sisson, I.,
 Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K.,
 Wu, C., Wu, X., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 226997)
 Worley, K.C.
 Direct Submission
 Submitted (24-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 ----- Genome Center of Medicine
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: KAMB
 Center clone name: CH230-80M23
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 185873 bases at least Q40
 Consensus quality: 190260 bases at least Q30
 Consensus quality: 193300 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 45 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 1681: contig of 1681 bp in length
 * 1682 1781: gap of unknown length
 * 1782 2911: contig of 1130 bp in length
 * 2912 3011: gap of unknown length
 * 3012 4486: contig of 1475 bp in length
 * 4487 4586: gap of unknown length
 * 4587 6300: contig of 1714 bp in length
 * 6301 7465: gap of unknown length
 * 7466 7565: contig of 1065 bp in length
 * 7566 9546: gap of unknown length
 * 9547 9646: contig of 1981 bp in length
 * 9647 11481: gap of unknown length
 * 11482 11581: contig of 1835 bp in length
 * 11582 13375: gap of unknown length
 * 13376 13475: contig of 1794 bp in length
 *
 * 13476 14903: contig of 1428 bp in length
 * 14904 15003: gap of unknown length
 * 15004 16512: contig of 1509 bp in length
 * 16513 16612: gap of unknown length
 * 16613 18713: contig of 2101 bp in length
 * 18714 20451: gap of unknown length
 * 20452 20551: contig of 1638 bp in length
 * 20552 22825: gap of unknown length
 * 22826 22825: contig of 2174 bp in length
 * 22827 24273: gap of unknown length
 * 24274 24373: contig of 1448 bp in length
 * 24374 26788: gap of unknown length
 * 26789 26888: contig of 2415 bp in length
 * 26889 29564: gap of unknown length
 * 29565 29564: contig of 2676 bp in length
 * 29566 32429: gap of unknown length
 * 32430 32529: contig of 2765 bp in length
 * 32530 35760: gap of unknown length
 * 35761 35860: contig of 3231 bp in length
 * 35861 38904: gap of unknown length
 * 38905 39004: contig of 3044 bp in length
 * 39005 43011: gap of unknown length
 * 43012 43131: contig of 4027 bp in length
 * 43132 46736: gap of unknown length
 * 46737 46836: contig of 3605 bp in length
 * 46837 49484: gap of unknown length
 * 49485 49584: contig of 2648 bp in length
 * 49585 54687: gap of unknown length
 * 54688 54787: contig of 5103 bp in length
 * 54788 58300: gap of unknown length
 * 58301 63148: contig of 3513 bp in length
 * 63149 63248: gap of unknown length
 * 63249 67905: contig of 4748 bp in length
 * 67906 68005: gap of unknown length
 * 68006 73317: contig of 4657 bp in length
 * 73318 73417: gap of unknown length
 * 73418 77498: contig of 5312 bp in length
 * 77499 77598: gap of unknown length
 * 77599 83850: contig of 4081 bp in length
 * 83851 83950: gap of unknown length
 * 83951 90459: contig of 6252 bp in length
 * 90460 90559: gap of unknown length
 * 90560 97151: contig of 6509 bp in length
 * 97152 97251: gap of unknown length
 * 97252 102654: contig of 6592 bp in length
 * 102655 102754: gap of unknown length
 * 102755 110357: contig of 5403 bp in length
 * 110358 110457: gap of unknown length
 * 110458 119224: contig of 7603 bp in length
 * 119225 119324: gap of unknown length
 * 119325 125038: contig of 8767 bp in length
 * 125039 125138: gap of unknown length
 * 125139 132952: contig of 5714 bp in length
 * 132953 133052: gap of unknown length
 * 133053 140429: contig of 7814 bp in length
 * 140430 140529: gap of unknown length
 * 140530 148095: contig of 7377 bp in length
 * 148096 148195: gap of unknown length
 * 148196 155089: contig of 7566 bp in length
 * 155090 155189: gap of unknown length
 * 155190 163298: contig of 6894 bp in length
 * 163299 163398: gap of unknown length
 * 163399 173674: contig of 8109 bp in length
 * 173675 173774: gap of unknown length
 * 173775 183596: contig of 10276 bp in length
 * 183597 183696: gap of unknown length
 * 183697 197884: contig of 9822 bp in length
 * 197885 197984: gap of unknown length
 * 197985 211084: contig of 14188 bp in length
 * 211085 211184: gap of unknown length
 * 211185 226997: contig of 13100 bp in length
 * 226998 226997: contig of 15813 bp in length.

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

FEATURES

Location/Qualifiers

source 1..226997
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-80M23"

BASE COUNT 66856 a 45772 c 45385 g 61745 t 7239 others
ORIGIN

Query Match

Best Local Similarity 7.1%; Score 210.2; DB 2; Length 226997;

Mismatches 0; Pred. No. 1.7e-44;

Matches 317; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

QY 868 TATCGGGCAGTGGGCGAGGGGTTCTACCTCGGACTGCCCACTGTCCAGAGCTAGAG 927

Db 158882 TCTGCGCGCGTACTTCGCGCGCGCGCCACTGCGACTGCCCACTGCCAGGAGCTGAG 158823

QY 928 CGGCTCGGGCAGCAGCGCGTGGCTGAGGAGAGAGCCCATTCACAGCTGCCACATCCCT 987

Db 158822 COTCTGGGCCAGCGGAGCGAGCGCTGCTGCGAAGGCTGCTAGCTGCCACATCCCA 158763

QY 988 GGGTCGGGAGGTGTAGCGCAAGCGCTTCGACTCTGAAGCCACTTCGCTGGCGACACT 1047

158762 GGCCTGGCGAAGGTGTAGCGCAAGCGCTTCGACTCTGAAGCGCGACCTTCGCTGGCATACC 158703

QY 1048 GCGGAGAGCGCTTCTGCTGCAACTGGCTTTCTCGCGCAGAGGTTCTCACTCGCTCTGAC 1107

Db 158702 GCGGAGCGCGCTTCTGCTGCAACTGGCTTTCTCGCGCAGAGGTTCTCACTCGCTCTGAC 158643

QY 1108 GAGCTGGAGCGCGCAGCTCGGCACTCACACCCGGGAGAGAGATTTCACCTTGCTGCTCTGT 1167

Db 158642 GAGCTGGAGCGCGCAGCTCGGCACTCACACCCGGGAGAGAGATTTCACCTTGCTGCTCTGT 158583

QY 1168 TCCAGCGCTTACAGAGAGCAGCAGCTTACAGCAACATCAGCGCACCAGCGGGAGCGCA 1227

Db 158582 AACAAAGCTTTCATCGCAGCGCAGCAGCAGCTTACAGCAACATCAGCGCAGCGGGTGGC 158523

QY 1228 GCGCGGGAGCGCGCGCGCGCTTAAAGAGCTGGGGAGGCTCGCAGGCTCGGGGAA 1287

Db 158522 GCGGCTCGCGGGCTCGGGTGGCGGCAAGAGGCGAGCGAGCGAGCGAGCGACAGC 158463

QY 1288 GAAGAAGCAATCAGCGCGCGCGCTTTCACCTTCGCTGCGCAGCGCGCGCGCGCGCGCGC 1347

Db 158462 GCGCGGGAAGCGCGCTTTCACCTTCGCTGCGCAGCGCGCGCGCGCGCGCGCGCGC 158403

QY 1348 GAGGAGCGCGCGAG 1362

Db 158402 CGCAATGCGCTGGAG 158388

RESULT 10

AC099580

INITIATION

AC099580 211696 bp DNA linear HTG 16-NOV-2001
Mus musculus clone RP23-119N8, WORKING DRAFT SEQUENCE, 22 unordered
pieces.

AC099580

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

1 (bases 1 to 211696)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguski, M., Bouck, J., Brown, A.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choquel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,

* 28332 31462: contig of 3131 bp in length
 * 31463 31562: gap of 100 bp
 * 31563 31952: contig of 3390 bp in length
 * 34953 35052: gap of 100 bp
 * 35053 41262: contig of 6210 bp in length
 * 41263 41362: gap of 100 bp
 * 41363 96204: contig of 54842 bp in length
 * 96205 96304: gap of 100 bp
 * 96305 113861: contig of 17557 bp in length
 * 113862 113961: gap of 100 bp
 * 113962 129625: contig of 15664 bp in length
 * 129626 129725: gap of 100 bp
 * 129726 144734: contig of 15009 bp in length
 * 144735 144834: gap of 100 bp
 * 144835 167414: contig of 22580 bp in length
 * 167415 167514: gap of 100 bp
 * 167515 205361: contig of 37847 bp in length
 * 205362 205461: gap of 100 bp
 * 205462 211696: contig of 6235 bp in length.

FEATURES

source

Location/Qualifiers
 1. .211696
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="RP23-119N8"
 /clone.lib="RPCI-23 Female Mouse BAC"
 1. .1138
 /note="assembly_fragment"
 1229. .3146
 /note="assembly_fragment"
 3247. .4633
 /note="assembly_fragment"
 4734. .6451
 /note="assembly_fragment"
 6552. .8169
 /note="assembly_fragment"
 8270. .9787
 /note="assembly_fragment"
 9888. .12231
 /note="assembly_fragment"
 12332. .14807
 /note="assembly_fragment"
 14908. .17935
 /note="assembly_fragment"
 18036. .21388
 /note="assembly_fragment"
 21489. .24719
 /note="assembly_fragment"
 24820. .28231
 /note="assembly_fragment"
 28332. .31462
 /note="assembly_fragment"
 31563. .34952
 /note="assembly_fragment"
 35053. .41262
 /note="assembly_fragment"
 41363. .96204
 /note="assembly_fragment"
 96305. .113861
 /note="assembly_fragment"
 113962. .129625
 /note="assembly_fragment"
 129726. .144734
 /note="assembly_fragment"
 144835. .167414
 /note="assembly_fragment"
 167515. .205361
 /note="assembly_fragment"
 205462. .211696
 /note="assembly_fragment"
 clone_end:T7
 vector_side:right"
 58957 a 45781 c 44552 g 50305 t 2101 others

Query Match 7.0%; Score 207.2; DB 2; Length 211696;
 Best Local Similarity 70.2%; Pred. No. 1.1e-43;
 Matches 278; Conservative 0; Mismatches 118; Indels 0; Gaps 0;
 QY 840 GGGTCTGGCACAGGGGCGGCTGGATATGCGGGCAGTGGGGCAGGCGTTCTACCTG 899
 Db 148978 GCGCGCGCGCGGCGCAGCTCGGCGCTCGCGCTACTCGGGCGCGCCACTG 149037
 QY 900 CGACTGCCCAACTGTACAGGAGCTAGAGCGGCTCGGGGCGAGCGGCTGGGCTGAGGAA 959
 Db 149038 CGACTGCCCAACTGTACAGGAGCGGCGGCTGGGCTGGGCGGCGGCGGCTGAGGAA 149097
 QY 960 GAAGCCCATTCACAGCTGCCACATCCCTGGGTGCGGCAAGGTGTACGGAAGGCTTCGCA 1019
 Db 149098 CAAGGGGCTGCACAGCTGTACATCCCGGGCTGTGCAAGGTGTACGGAAGGCTTCGCA 149157
 QY 1020 TCTGAAGCCCACTTGGCGCTGGCACATGGCGAGAGGCTTTCTGCTGCAACTGGCTTTT 1079
 Db 149158 CCTGAAGGGCGACCTTGGCGCTGGCACACAGGGGCGGCGGCTTCTGCTGCAACTGGCTTT 149217
 QY 1080 CTGCGGCAAGAGTTCACTCGCTCTGACGAGCTGGAGCGCCACGCTGCGCACTCACACCGG 1139
 Db 149218 CTGCGGCAAGAGTTCACTCGCTCTGACGAGCTGGAGCGCCACGCTGCGCACTCACACCGG 149277
 QY 1140 GGAGAAGAGTTCACTCGCTCTGTTTCCAAAGCGCTTTACAGAGCGGCGGCTGAG 1199
 Db 149278 CGAGAAGCGCTTGGCGCTGCTGTTGCAACAAGCGTTTTCATGCGGCGGCGGCGGCTGAG 149337
 QY 1200 CAACATCAGCGCCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1235
 Db 149338 CAACACATTAAGACGCAACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 149373

RESULT 11

AC126277/c 244434 bp DNA linear HTG 04-JUL-2002
 LOCUS Mus musculus chromosome UNK clone RP23-161L22, WORKING DRAFT
 DEFINITION SEQUENCE, 5 unordered pieces.
 ACCESSION AC126277
 VERSION AC126277.1 GI:21694006
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 McPherson, J.D. and Waterston, R.H.
 The sequence of Mus musculus clone
 Unpublished
 McPherson, J.D. and Waterston, R.H.
 Direct Submission
 Submitted (04-JUL-2002) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA

COMMENT

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc/index.shtml>
 Contact: submissions@watson.wustl.edu
 ----- Project Information -----
 Center project name: M.BA0161L22
 ----- Summary Statistics -----
 Sequencing vector: M13; 0%
 Sequencing vector: plasmid; 100%
 Chemistry: Dye-terminator; 0% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 242117 bases at least Q40
 Consensus quality: 242702 bases at least Q30
 Consensus quality: 242985 bases at least Q20

AGCCCAATCAGCCGCCCGATCTTCCACTTCGCGCTGCAC

[illegible]


```

Db 23198 CCGCACTGGCAGAGGTTGAGCTGTGCCCCACCGGAGCGAGCTGCGTCGCAAGG 23139
QY 966 CATTACAGCTGCCACATCCCTGGGTGGGCAAGGTGTACGGCAAGGTTCCGCTATCGAA 1025
Db 23138 GCTGCATAGCTGCCACATCCCGAGGCTGGGCAAGGTGTACGGCAAGGTTCCGACCTGAA 23079
QY 1026 ACCGACCTGGCTGGGACATCGGGAGAGGCTTTCGTCTGCAACTGGCTTTCTCGGG 1085
Db 23078 GCGCACCCTGGCTGGCATCGGAGCGGCGCTTTCGTCTGCAACTGGCTTTCTCGGG 23019
QY 1086 CAAGAGGTTCACTCGCTGTGACGAGCTGGAGCGCCACGTCGCGCACTACACCCCGGAGAA 1145
Db 23018 TAAGGCTTTAGCGCTCCGACGAGCTGCAGCGCACTTCGGACCCACACGGCGGAGAA 22959
QY 1146 GAAGTTCACTGCTGCTCTGTTCCAGCGCTTTACCAAGACGACCACTTGAGCAACA 1205
Db 22958 GCGCTTCGCTGCGCGGTGTGCAACAACGCTTCATGCGCAGCGACCACTTGAGCAACA 22899
QY 1206 TCAGCGCACCGGAGCGGAGCGGCGGAGCGGCGGAGCGGCGGAGCGGCGGAGCGGAG 1265
Db 22898 CGTGAAGACACTACGCGGCGGAGCGGCGGAGCGGCGGAGCGGCGGAGCGGAGCGGAG 22839
1266 GGAGGCTCGCAGCGTCGCGGGAAGAGCAATCAGCGCGCGCGGATC 1313
Db 22838 CGACACCCACATCGAGGACGCGGCGGAGCGGCGGAGCGGCGGAGCGGCGGAGCGGATC 22791

```

RESULT 13

```

AC014235/c
LOCUS AC014235 41507 bp DNA linear HTG 16-NOV-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC014235
VERSION AC014235.1 GI:6437100
KEYWORDS HTG; HTGS PHASE2.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 41507)
AUTHORS Adams,M. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10211572 by the submitter.
For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

```

.TUES

```

source
1. 41507
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 12153 a 8618 c 8808 g 11928 t
ORIGIN

```

Query Match

```

Best Local Similarity 6.1%; Score 180.4; DB 2; Length 41507;
Matches 265; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

```

```

QY 866 GATATCGGGCAGTGGGCGGCGGTTCTACCTGCGACTGCCCACTGTCCAGGAGTAG 925
Db 30339 GATCCAGAGGGGTTATCCGCGGCGGACATGCTGATTCGCGAATTCGCGAGGGCGG 30280
QY 926 ACCGCTCGGGCAGCAGCGGCTGGGTGAGGAGAACCCCATTCACAGCTGCCACATCC 985
Db 30279 AACGCTCGGGTCCCGCTGGCTTCATCTGCGCAAAAAGAACATCCATCGTGTACATAC 30220
QY 986 CTGGGTGCGGCAAGGTGTACGCAAGGCTTCGCATCTCTAAAGCCGACCTGCGCTGGGACA 1045

```

```

Db 30219 CTGCTCTGGCAAGGTGTATGGAAGAGCTCACATCTGAAGGCTCACCTGCGGTGGCACA 30160
QY 1046 CTGCGAGAGAGCTTTTCTGCTGCAACTGGCTTTTCTGCGCAAGAGGTTCTACTGCTCTG 1105
Db 30159 CCGCGAGAGGCTCCCTTGTGTGCAACTGGCTTGTCTGCGCAAGGTTTTCACAGCTCCG 30100
QY 1106 ACAGAGCTGGAGCGGACGCTGCGCACTACACCCGGGAGAGAAAGTTTCACTTGGCTGCTCT 1165
Db 30099 ACAGAGCTGCAAGGCACTTTGAGAGCGCACACCGGCGGAGAGGCTTTCATGCTCCCGGT 30040
QY 1166 GTTCCAGAGCGCTTTACAGAGCGACCACTTTGAGCAACATCAGCGCACCCACGGGAGC 1225
Db 30039 GCAATAGAGGCTTTATGCGCAGCGACCATTTGGCCACACGCTGAAGACGCAATAGGCA 29980
QY 1226 CAGCGCCGGGACCGCCCAAGTGGCCCTAAGGAGCTGGGGAGGG 1271
Db 29979 CGGCCAATCACCAGGCCAATGGCCACATGCTGGGCTGAAGAAGGG 29934

```

RESULT 14

```

AC023702 159970 bp DNA linear INV 13-JUN-2002
LOCUS AC023702
DEFINITION Drosophila melanogaster X BAC RP98-24L2 (Roswell Park Cancer
Institute Drosophila BAC Library): complete sequence.
ACCESSION AC023702
VERSION AC023702.4 GI:21397187
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 159970)
AUTHORS Muzny,D., Scherer,S., Adams,M.D., Holt,R.A., Evans,C.A.,
Gocayne,J.D., Tabor,P., Williamson,A., Homsí,F.H.,
Dugan-Rocha,S.D., Sodergren,E.S., Hodgson,A.H., Chen,R.C.,
Ayele,M., Scott,G.S., Worley,K.W., Ananides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Balwin,D., Beeson,K.Y., Brown,M., Buhay,C.,
Busam,D.A., Center,A., Chen,G., Chen,Z., Clerc-Blankenburg,K.,
Davenport,L.B., Dietz,S.M., Ding,Y., Dodson,K., Doup,L.E.,
Draper,H., Emery-Cohen,A., Ferreira,S., Garg,N.D.S., Houck,J.,
Hostin,D., Howland,T.J., Hume,J., Ibegwam,C., Jalali,M., Kovar,C.,
Liu,W., Mattei,B., McIntosh,T.C., Morgan,M., Moy,M., Murphy,B.,
Neelson,K.A., Ndassa,Y., Nguyen,N., Perez,L., Pittman,G.S., Puri,V.,
Schaefer,F., Shen,H., Strong,R., Tector,C., Wang,Q., Williams,S.M.,
Xiang,J., Zaveri,J.S., Zhou,J., Zorrilla,S., Smith,H.O.,
Wheeler,D., Weinstock,G., Gibbs,R. and Venter,J.C.
Direct Submission
Unpublished
2 (bases 1 to 159970)

```

TITLE

```

JOURNAL
AUTHORS
Worley,K.C., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowen,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,S., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,
Hollins,B., Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,D.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratochvil,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulseghed,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,

```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2003, 14:15:28 ; Search time 404 Seconds
(without alignments)
16499.790 Million cell updates/sec

Title: US-09-734-329-1
Perfect score: 2960
Sequence: 1 attctccattctccctccc.....aaaaaaaaaaaaaaaaaaaaa 2960

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	339.2	11.5	1025	20	AA330319
2	326.2	11.0	1065	20	AA330407
3	180.4	6.1	2076	23	ABL02405
4	180.4	6.1	4311	23	ABL02404
5	171.6	5.8	1675	23	ABL079321
6	159.6	5.4	1830	24	ABA96948
7	159	5.4	1197	24	ABA96949
8	141.8	4.8	882	22	AA526463
9	141.8	4.8	882	22	AA558802
					DNA encoding a hum
					DNA encoding a hum
					Drosophila melanog
					Drosophila melanog
					DNA encoding novel
					Mouse Spl family t
					Human Spl family t
					Human cDNA encodin
					Human immune/haema

10	141.8	4.8	900	22	AA526032	Human cDNA encodin
11	141.8	4.8	1430	21	AAA50460	Human zinc finger
12	136.8	4.6	1471	21	AAA50468	Mouse zinc finger
13	134.6	4.5	3090	22	AAI60335	Human polynucleoti
14	134.6	4.5	3289	22	AAI58549	Human polynucleoti
15	130.8	4.4	1615	24	ABQ54824	Human ovarian anti
16	130.8	4.4	3288	24	ABQ54824	Human ovarian anti
17	130.8	4.4	4285	18	ABK92438	Human cDNA differe
18	127	4.3	894	18	AA749492	Human prostate spe
19	126.8	4.3	912	18	AA749493	Chimeric restricti
20	125.6	4.2	298	21	AA59164	Chimeric restricti
21	125.6	4.2	298	22	AAI5335	DNA encoding zinc
22	125.6	4.2	298	22	AAI5335	Human ZFP-vascular
23	124	4.2	298	21	AAI67953	Human Veg 1 domain
24	124	4.2	298	22	AAI5336	DNA encoding zinc
25	124	4.2	298	24	AAI67956	Human ZFP-vascular
26	122.4	4.1	589	21	AA59176	Human Veg 3a domai
27	122.4	4.1	589	22	AAI5346	DNA encoding zinc
28	122.4	4.1	589	24	AAI67966	Human ZFP-vascular
29	122.4	4.1	3187	23	ABL09749	6-finger zing fing
30	122.4	4.1	3388	23	ABL10734	Drosophila melanog
31	122.4	4.1	5187	23	ABL09748	Drosophila melanog
32	120	4.1	292	22	AA57392	Drosophila melanog
33	118.8	4.0	3548	24	ABK84634	DNA encoding zinc
34	112.6	3.8	4859	24	ABN95149	Human cDNA differe
35	112.6	3.8	4859	24	ABL65183	Gene #1647 used to
36	110.2	3.7	667	24	ABQ17940	Lung cancer relate
37	110.2	3.7	667	24	ABQ17941	Oligonucleotide fo
38	109.6	3.7	3069	24	ABI99387	Oligonucleotide fo
39	109.2	3.7	1935	23	ABL27239	Mouse ischaemic co
40	109.2	3.7	4374	23	ABL27238	Drosophila melanog
41	108.6	3.6	547	24	ABL81965	Drosophila melanog
42	107	3.6	1580	23	ABL05599	Human ovarian canc
43	107	3.6	2881	18	AA72160	Drosophila melanog
44	107	3.6	2909	23	AA573995	Transforming growt
45	107	3.6	3217	22	AAH02903	DNA encoding novel
						Human shear stress

ALIGNMENTS

RESULT 1
AA330319/c
ID AAX30319 standard; DNA; 1025 BP.

XX AAX30319;

XX 14-MAY-1999 (first entry)

XX DNA encoding a human secreted protein.

XX Secreted protein; cancer; tumour; neurodegenerative disorder;
XX developmental abnormality; foetal deficiency; blood disorder;
XX CNS disorder; immune system disease; autoimmune disease; hepatic disease;
XX renal disease; diabetes; inflammation; allergy; ischemic shock;
XX Alzheimer's; cognitive disorder; schizophrenia; cardiovascular disorder;
XX prostate disease; asthma; osteoporosis; arthritis; ss.

OS Homo sapiens.

XX WO9907891-A1.

XX 18-FEB-1999.

XX 04-AUG-1998; 98WO-US16235.

XX 19-AUG-1997; 97US-0056732.

XX 05-AUG-1997; 97US-0054798.

XX 05-AUG-1997; 97US-0054803.

XX 05-AUG-1997; 97US-0054804.

XX 05-AUG-1997; 97US-0054806.

XX 05-AUG-1997; 97US-0054807.

XX 05-AUG-1997; 97US-0054808.

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 1697.
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-658860/75.
DR P-PSDB; ABB58302.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
XX Claim 1; SEQ ID NO 1697; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 2076 BP; 418 A; 664 C; 629 G; 365 T; 0 other;

Query Match 6.1%; Score 180.4; DB 23; Length 2076;
Best Local Similarity 65.3%; Pred. No. 4.8e-32;
Matches 265; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
QY 866 GATATCGGGCAGTGGGCGGCTTCTACCTGCGACTGCCCACTGTCCAGGAGCTAG 925
DB 977 GATCCAGAGGCGTTATCGCGAGGGCCACATGCGATTCGCCGATTCAGGAGCGG 1036
926 AGCGGCTCGGGCAGGCGGCTGGGCTGAGGAAGGCCATTCACAGTCCCATCC 985
UB 1037 AACGCTGGTCCGCTGGCTTCATCTCGCAAAAAGAACATCCATTCGTGTCACATAC 1096
QY 986 CTGGGTGCGCAAGGTGTACGCAAGGCTTCGATCTGAAGCCCACTTGCGCTGCACA 1045
DB 1097 CTGGGTGCGCAAGGTGTATGAAGACGTCATCTGAAGGCTCACCTCGGCTGCACA 1156
QY 1046 CTGGCGAGAGGCTTTCGTCTGCAACTGGCTTTCTGCGGCAAGAGGTTCACTCGCTCG 1105
DB 1157 CCGCGCAGGCTCCCTTTGTGTGCAACTGGCTGTCTGCGCAAGGCTTTCACAGCTCG 1216
QY 1106 ACAGCTGAGCGCCAGCTGCGCACTACACCGGGGAGAGGTTCACTTCGCTGCT 1165
DB 1217 ACAGCTGCAACGGCATTTGAGGACGCACACCGGCGAGAAGGCTTTGCTATGCCCGTGT 1276
QY 1166 GTTCCAGGCTTTTACCAAGAGGACCACTTGAGCAACATCAGCCGCCACCGGGGAGC 1225
DB 1277 GCAATAGGCTTTATGCGCAGGCGCCATTTGGCCCAAGCAGCTGAAGACGCAATGCA 1336
QY 1226 CAGGCCCGGAGCAGCGCCCAAGTGGCCCTAAGGAGCTGGGGAGGG 1271

DB 1337 CGGCCAATCACCAGCCCAATGCCACATGTGGCTGAAGAGGG 1382
RESULT 4
ABL02404
ID ABL02404 standard; cDNA; 4311 BP.
XX ABL02404;
AC ABL02404;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 1694.
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-658860/75.
DR P-PSDB; ABB58301.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
XX Claim 1; SEQ ID NO 1694; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 4311 BP; 1189 A; 1068 C; 1052 G; 1002 T; 0 other;
Query Match 6.1%; Score 180.4; DB 23; Length 4311;
Best Local Similarity 65.3%; Pred. No. 6.3e-32;
Matches 265; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
QY 866 GATATCGGGCAGTGGGCGGCTTCTACCTGCGACTGCCCACTGTCCAGGAGCTAG 925
DB 2212 GATCCAGAGGCGTTATCGCGAGGGCCACATGCGATTCGCCGATTCAGGAGCGG 2271
QY 926 AGCGGCTCGGGCAGCAGCGGCTGGGCTGAGGAAGGCCATTCACAGTCCCATCC 985
DB 2272 AACGCTGGTCCGCTGGCTTCATCTCGCAAAAAGAACATCCATTCGTGTCACATAC 2331
QY 986 CTGGGTGCGCAAGGTGTACGCAAGGCTTCGATCTGAAGCCCACTTGCGCTGCACA 1045
DB 2332 CTGGGTGCGCAAGGTGTATGAAGACGTCATCTTGAAGGCTCACCTCGGCTGCACA 2391
QY 1046 CTGGGCGAGGCGCTTTCGTCTGCAACTGGCTTTCTGCGGCAAGAGGTTCACTCGCTCG 1105
DB 2392 CCGGCGAGGCTCCCTTTGTGTGCAACTGGCTTTCTGCGGCAAGGCTTTCACACGCTCG 2451


```
XX PR 20-JUN-2000; 2000JP-0189762.
XX PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX PI Senoo C, Numata M;
XX XX WPI: 2002-114585/15.
DR DR P-PSDB; AAM49116.
XX XX
XX XX Spl family transcription factor having zinc finger domain for use as a
PT PT target for drug development
XX XX Claim 1; Fig 1; 96pp; Japanese.
XX XX
XX XX This sequence represents cDNA encoding a murine Spl family transcription
CC CC factor, m285, and its human homologue, h285 (AAM49117) contain 3
CC CC zinc finger domains and are present in the nucleus, where they bind GC
CC CC boxes and regulate transcriptional activity. The invention also
CC CC encompasses fragments of the 285 proteins and nucleic acids encoding
CC CC them, expression vectors and host cells comprising 285 protein-encoding
CC CC nucleic acids, the recombinant production of 285 proteins, antibodies
CC CC against 285 proteins, a method of screening for compounds which bind to
CC CC 285 proteins, and a method for the regulation of the transcription
CC CC activity of promoters, especially of the CMV promoter, using the 285
CC CC protein. The 285 proteins can be used target molecules for drug
XX XX development and for gene therapy.
XX XX
XX XX Sequence 1830 BP; 349 A; 639 C; 514 G; 326 T; 2 other;
SQ
Query Match 5.4%; Score 159.6; DB 24; Length 1830;
Best Local Similarity 68.4%; Pred. No. 3.3e-27;
Matches 238; Conservative 0; Mismatches 104; Indels 6; Gaps 1;
QY 871 GCGGGCAGTGGGCGAGGGCGTTCTACTGCGACTGCCCAACTGCGAGGCTAGAGCGG 930
DB 867 GCGGCCACGCCAGGAGGTGCCCGCTGCGGTGCCCAACTGCCAGGGCTG----- 921
QY 931 CTGGGGCAGCAGCGGTGGGTGAGGAAGAAGCCCATTCACAGCTGCCACATCCCTGGG 990
DB 922 -GCGGTGCCCGGAGGCGGAACCGGGCAAAAGAAAGCAACACGTGTGCCACGTGCCAGG 980
QY 991 TGGGGCAAGGTGTACGCGAAGGCTTGCATCTCTGAAGCCCACTGCGTGGCACATGCG 1050
DB 981 TGTGGCAAGGTGTACGCGAAGGCTTGCATCTCTGAAGCCCACTGCGTGGCACATGCG 1040
QY 1051 GAGAGGCGTTTCTGTCGAACCTGGCTTTCTGCGGCAAGAGTTTCACTGCTGTACGAG 1110
DB 1041 GAGCGGCCCTTCGTGTGCAACTGGCTTCTTCTGCGGCAAGAGTTTCACTGCTGTACG 1100
QY 1111 CTGGAGCGCCACGTGGCGACTACACCGGGGAGAAAGTTTCACTTGCCTGTCTCTTCC 1170
DB 1101 CTGCAAGGCGACCTGGGGAGTACACGGGGGAGAAAGGCTTCTGCTTGTCCGAGTGGCG 1160
QY 1171 AAGCGCTTTACCAAGAGCGACCACTTGAGCAACATCAGCGCACCCAC 1218
DB 1161 AAACGCTTTCAGGAGCGATACCTCGGCCAAGCAGCTGAAGACGCAC 1208
RESULT 7
ID ABA96949
XX AC ABA96949;
XX XX
XX XX 20-MAY-2002 (first entry)
XX XX Human Spl family transcription factor h285-encoding cDNA.
XX XX Human; h285; transcription factor; Spl family; zinc finger;
KW GC-box binding; 285 protein; promoter activity regulation;
KW transcriptional regulation; drug screening; drug development;
KW gene therapy; gene; ss.
```

```
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT 1..1197
XX FT /*tag= a
XX FT /product= "h285"
XX FT /function= "transcription factor"
XX FT /note= "Spl family; binds GC-boxes"
XX PN WO200198492-A1.
XX XX
XX PD 27-DEC-2001.
XX PF 14-JUN-2001; 2001WO-JP05066.
XX PR 20-JUN-2000; 2000JP-0189762.
XX XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX PI Senoo C, Numata M;
XX DR WPI: 2002-114585/15.
DR DR P-PSDB; AAM49117.
XX XX Spl family transcription factor having zinc finger domain for use as a
PT PT target for drug development
XX XX Claim 1; Fig 6; 96pp; Japanese.
XX XX
XX XX This sequence represents cDNA encoding a human Spl family transcription
CC CC factor, h285, and its murine homologue, m285 (AAM49116) contain 3
CC CC zinc finger domains and are present in the nucleus, where they bind GC
CC CC boxes and regulate transcriptional activity. The invention also
CC CC encompasses fragments of the 285 proteins and nucleic acids encoding
CC CC them, expression vectors and host cells comprising 285 protein-encoding
CC CC nucleic acids, the recombinant production of 285 proteins, antibodies
CC CC against 285 proteins, a method of screening for compounds which bind to
CC CC 285 proteins, and a method for the regulation of the transcription
CC CC activity of promoters, especially of the CMV promoter, using the 285
CC CC protein. The 285 proteins can be used target molecules for drug
XX XX development and for gene therapy.
XX XX
XX XX Sequence 1197 BP; 169 A; 486 C; 385 G; 157 T; 0 other;
SQ
Query Match 5.4%; Score 159; DB 24; Length 1197;
Best Local Similarity 67.9%; Pred. No. 3.9e-27;
Matches 222; Conservative 0; Mismatches 105; Indels 0; Gaps 0;
QY 897 CTGCGACTGCCCAACTGTCTGAGGAGCTAGAGCGGCTCGGGGAGAGCGGCTGGGCTGAG 956
DB 813 CGCGCGCTCGCGTGTCCCAACTGCCAGCGCGGCGCGCGCGCGCGAGCGGCGGCGG 872
QY 957 GAAGAAGCCCATTCACAGCTGCCACATCCCTGGTGGTGGGAGAGGCTGTAGCGAAGCTTC 1016
DB 873 GAAGAAGAAGCAGCAGCTGTGCCACGTGCGGCGCTCGCGAGGTGTACGGGAAGACGTC 932
QY 1017 GCATCTGAAAGGCCCACTTGGCTGGCACACTGGCGAGAGGCGCTTCTGCTGCAACTGGCT 1076
DB 933 GCACCTGAAGGGCGACCTCGCTGGCACAGCGGCGAGCGACCCCTTCTGTGCAACTGGCT 992
QY 1077 TTTCTGCGGCAAGAGGTTTCACTTCGCTCTGACGAGCTGGAGCGGCGAGCTGCACAC 1136
DB 993 CTTCTGCGGGAAGAGTTTCACTTCACGCGCTCGGACGAGCTGCAGGCGACCTGCGGACTCACAC 1052
QY 1137 CCGGAGAGAGAGTTTCACTTGGCTGTCTGTCTTCCAAAGCGCTTTACAGAGCGACCACTT 1196
DB 1053 GGGCGAGAAAGCGCTTGTGCTGTCCGAGTGGCGAAGCGCTTCATCGCAGCGACCACT 1112
QY 1197 GAGCAAAATCATCGCGCACCCACCGGGA 1223
DB 1113 CGGGAAGCAGCTCAAGACTCACAGAA 1139
```

RESULT 8
AAS26463
ID AAS26463 standard; cDNA; 882 BP.
XX AC AAS26463;
XX DT 07-NOV-2001 (first entry)
XX DE Human cDNA encoding a novel secreted protein, Seq ID 642.
XX KW Human; immunosuppressive; antiarthritic; ss; antirheumatic;
KW cytosolic; cardiac; vasotropic; cerebroprotective; nootropic;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW vulnary; secreted protein; rheumatoid arthritis;
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW corneal infection; wound healing; epithelial cell proliferation;
KW skin ageing; food additive; preservative; antiproliferative.
OS Homo sapiens.
XX WO200155322-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01341.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 25-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 29-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.

PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
PR XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI: 2001-488783/53.
DR P-PSDB; AAU16476.
XX
PT New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives.
XX
XX Claim 1: SEQ ID No 642; 980pp; English.
XX
XX The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility
CC to a pathological condition. Antibodies to the proteins can also
CC be used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection, and many other
CC disorders listed in the specification. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence encodes a novel secreted protein of the invention.

Query Match 4.8%; Score 141.8; DB 22; Length 882;
Best Local Similarity 69.0%; Pred. No. 3.7e-23;
Matches 194; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 960 GAAGCCCAATTCAGCTGCCACATCCCTGGTGGCGCAAGGTGTACGGCAAGCTTCGCA 1019
DB 132 GCAGAGGAGACCAAGTGCCTACCTACCGGCTGCCGAGAAATTCGGAATCTCGCA 191
QY 1020 TGTGAAGCCCACTTGGCTGCCACACTGGCGAGAGGCTTTCGCTGCAACTGCGCTTT 1079
DB 192 CTTCAAGCGGCACCTGAGAACTACACAGGTGAGAGGCGCTTCGCGCTGCGAGTCGACGA 251
QY 1080 CTGCGGCAAGAGGTTCACCTGCTGACGAGCTGAGCGGCCACTGCGCACTCACACCG 1139
DB 252 CTGCAACAAGAAGTTGGCGCGCTCCGACGAGCTGGCGGGGCACTACCGCACACACCGG 311

QY 1140 GGAGAGAAGTTCACTTCTGCTCTGTTCCAAAGCGCTTTACGAGAAGCGACCACTTGAG 1199
DB 312 CGAGAAGAAGTTTCAGTCCCATCTGCGAGAAGCGCTTCATGCGAGCGACCACTGAC 371
QY 1200 CAAACATCAGCGCACCCAGCGGGGAGCGCCGCGGACCGC 1240
DB 372 CAAGCACGCGCGCGCCAGCAACTTCCACCCCGGAATGC 412

RESULT 9
AAK58802
ID AAK58802 standard; cDNA; 882 BP.
XX
AC AAK58802;
XX
DT 06-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:3862.
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ss.
XX Homo sapiens.
XX WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 26-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.


```
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246533.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.

PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Barash SC, Ruben SM;
XX DR WPI; 2001-483426/52.
XX DR P-PSDB; AAM86021.
XX PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX PT useful for preventing, diagnosing and/or treating cancers and
XX PT metastasis.
XX PS Claim 1; SEQ ID NO 3862; 3071pp + Sequence Listing; English.
XX CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
XX CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
XX CC activity, and can be used in gene therapy and vaccine production. (I)
XX CC proteins and polynucleotides may be used in the prevention, diagnosis and
XX CC treatment of diseases associated with inappropriate (I) expression. For
XX CC example, they may be used to treat disorders associated with decreased
XX CC expression by rectifying mutations or deletions in a patient's genome
XX CC that affect the activity of (I) by expressing inactive proteins or to
XX CC supplement the patients own production of (I). Additionally, (I)
XX CC polynucleotides may be used to produce the secreted (I), by inserting
XX CC the nucleic acids into a host cell and culturing the cell to express the
XX CC protein. (I) proteins and polynucleotides may be used to prevent
XX CC cancers and cancer metastases of haematopoietic-related diseases, especially
XX CC to AAK87694 represent human immune/haematopoietic antigen genomic
XX CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
XX CC represent sequences used in the exemplification of the present invention.
XX SQ Sequence 882 BP; 221 A; 279 C; 240 G; 137 T; 5 other;

Query Match 4.8%; Score 141.8; DB 22; Length 882;
Best Local Similarity 69.0%; Pred. No. 3.7e-23;
Matches 194; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 960 GAAGCCCATTCACAGCTGCCACATCCCTGGTGGGGGAGAGGCTTTCGCTGCACTGCTGCTTTCGCA 1019
Db 132 GCAGAGGAAGCAAGTCCCACTAGCGGGCTGCGAGAAAGTTTACGGGAATCTTCGCA 191
QY 1020 TCTGAAAGCCCATTCGCTGCGCTGCGACACTGCGAGAGGAACTTTCGCTGCACTGCTGCTT 1079
Db 192 CCTAAGGCGCCTGAGAACTCACAAGTGCAGAGGCGCTTCGCTGCACTGCGACAGA 251
QY 1080 CTGCGGCAAGAGTTTCACTGCTCTGACAGCTGAGAGCGCCACGCTGCGCACTCACACCG 1139
Db 252 CTGCAACAAGAAGTTGCGCGCTCCGACGAGCTGCGCGGCGCACTACCGCACACACACGCG 311
QY 1140 GGAGAAGAGTTCACTTGCCTGCTGCTGTTCCAGAGGCTTTTACCAGAGGACCACTTACG 1199
|||||
```

Db 312 CGAAGAGATTTCAGTCCGCTTCGAGAGCGCTTCATGCGCAGGACCACTGCAC 371
QY 1200 CAAACATCAGCCAGCCAGCGGAGCGAGCGCCGCGGACCGC 1240
Db 372 CAAGCAGCGCGCGCCGCGCAGCGCAACTTCACCCGGGAATGC 412

RESULT 10
AAS26032
ID AAS26032 standard; cDNA; 900 BP.
XX
AC AAS26032;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human cDNA encoding a novel secreted protein, Seq ID 211.
XX
KW Human; immunosuppressive; antiarthritic; ss; antirheumatic;
KW cytotatic; cardiant; vasotropic; cerebroprotective; nootropic;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW vulnary; secreted protein; rheumatoid arthritis;
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;
nervous system disorder; Alzheimer's disease; infection; ocular disorder;
corneal infection; wound healing; epithelial cell proliferation;
skin ageing; food additive; preservative; antiproliferative.
OS Homo sapiens.
XX
PN WO200155322-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01341.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214880.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.

PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249246.
 PR 17-NOV-2000; 2000US-0249285.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250321.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Barash SC, Ruben SM;
 PI WPI: 2001-488783/53.
 DR P-PSDB; RAU16045.
 XX
 PT New nucleic acid molecules encoding 461 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives.
 XX Claim 1; SEQ ID NO 211; 980pp; English.
 XX
 CC The invention relates to isolated nucleic acid molecules and their
 CC encoded secreted proteins. The nucleic acids and proteins are used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
 CC are also used in diagnosing a pathological condition or susceptibility
 CC to a pathological condition. Antibodies to the proteins can also
 CC be used in alleviating symptoms associated with the disorders and in
 CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
 CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated
 CC include autoimmune diseases e.g. rheumatoid arthritis,
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
 CC e.g. cerebral ischaemia, angioneurosis, nervous system disorders e.g.
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
 CC and ocular disorders e.g. corneal infection, and many other
 CC disorders listed in the specification. The polypeptides can also
 CC be used to aid wound healing and epithelial cell proliferation, to
 CC prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. The present
 CC sequence encodes a novel secreted protein of the invention.
 Query Match 4.8%; Score 141.8; DB 22; Length 900;
 Best Local Similarity 69.0%; Pred. No. 3.7e-23;
 Matches 194; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
 QY 960 GAAGCCCATTCACAGCTGCCACATCCCTGGGTGGCGGCAAGGTGTACGGCAGGCTTCGCA 1019
 Db 150 GCAGAGGAAGCACAAGTGCCTACTACGGGGCTGGGAGAAAGTTTACGGGAATCTTCGCA 209
 QY 1020 TCTGAAGCCCATTCGGCTGGCAGACACTGGCGAGAGGCCCTTCTGCTGCAACTGGCTTT 1079

Db 210 CCTCAAGCGCGACCTGAGAACTCACACAGTGTGAGAGGCCCTTCGCTGAGTGCAGGA 269
 QY 1080 CTGCGGCAAGAGTTTCACTCGCTCTGAGAGTGTGAGCGCCGACGTGCGCACTCACACCG 1139
 Db 270 CTGCAACAGAAGTTGCGCGCTCGGAGAGTGTGCGCGGCACTACCGCACACACCGG 329
 QY 1140 GGAGAAGAAGTTCACTTGCCTGCTCTGTTCGAAGCGCTTACCAAGAAGCGACCTTGAG 1199
 Db 330 CGAGAAGAAGTTCACTTGCCTGCTCGGAGAGCGCTTTCATGCGCAGGACCACTGAC 389
 QY 1200 CAACATCAGCGCACCGAGCGGAGCGGAGCGGCGGAGCGG 1240
 Db 390 CAAGCAGCGCGCGCGCACGCAACTTCCACCGCGGAATGC 430
 RESULT 11
 AA50460
 ID AA50460 standard; cDNA; 1430 BP.
 XX AC
 XX AAA50460;
 XX
 DT 05-DEC-2000 (first entry)
 XX
 DE Human zinc finger transcription factor RFLAT-1 cDNA.
 XX
 KW RFLAT-1; RANTES factor of late activated T-lymphocyte;
 KW transcription factor; zinc finger; human; inflammation;
 KW antiinflammatory; autoimmune disease; AIDS; neoplasia;
 KW gene therapy; diagnosis; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 379..1245
 FT /*tag= a
 FT /transl_except= (pos:1225..1227, aa:Thr)
 FT /misc_feature 370..381
 FT /*tag= b
 FT /note= "consensus Kozak sequence"
 XX
 XX WO200052030-A1.
 XX
 XX 08-SEP-2000.
 XX
 XX 27-JAN-2000; 2000WO-US02300.
 XX
 XX 27-JAN-1999; 99US-0117576.
 XX
 XX (STRD) UNIV LELAND STANFORD JUNIOR.
 XX
 XX Song AM, Chen Y, Krensky AM;
 XX
 XX WPI: 2000-587291/55.
 XX P-PSDB; AAY95986.
 XX
 XX New regulated upon activation, normal T cell expressed and secreted
 XX (RANTES) factor of late activated T lymphocytes-1 proteins and nucleic
 XX acids, for treating conditions associated with RANTES activity, e.g.
 XX inflammation, AIDS
 XX
 XX Claim 5; Fig 1; 53pp; English.
 XX
 XX This is the nucleotide sequence of a 1430 bp cDNA clone encoding
 XX novel human RFLAT-1 (regulated upon activation, normal T cell
 XX expressed and secreted (RANTES) factor of late activated T
 XX lymphocytes-1). To isolate RFLAT-1 cDNA, an oligonucleotide probe
 XX (AA50461) was used to screen a lambda-gt11 cDNA library constructed
 XX using polyA+ RNA from PHA-stimulated human peripheral blood cells.
 XX A cDNA clone (980 bp) was isolated, and was used to rescreen the
 XX library to identify the 1430 bp clone. RFLAT-1 (see AAY95986) is a
 XX novel zinc finger transcription factor that activates RANTES gene
 XX expression by binding to the A region of the RANTES gene promoter.

CC RFLAT-1 polypeptides and nucleic acids are used in diagnostic and
CC therapeutic agent screening, discovery, and preparation applications,
CC as well as in modulating RANTES activity, especially in the treatment
CC of conditions associated with RANTES activity, such as in the
CC suppression of RANTES gene expression, e.g. acute or chronic
CC inflammation, autoimmune diseases, and in conditions in which
CC enhancement of RANTES gene expression is needed, e.g. AIDS, or
CC neoplastic diseases. The RFLAT-1 gene, gene fragments, or the
CC encoded protein are useful in gene therapy to treat disorders
CC associated with gene defects or disease conditions in which
CC enhanced gene activity is desired. The gene may also be used in
CC the identification of RFLAT-1 gene homologues, as a source of novel
CC promoter elements, in the identification of RFLAT-1 expression
CC regulatory factors, to produce as probes and primers, in the
CC identification of expression patterns in a biological specimens,
CC and in the preparation of cell or transgenic animal models of
CC RFLAT-1 protein function.
XX
SQ Sequence 1430 BP; 250 A; 550 C; 472 G; 158 T; 0 other;

Query Match 4.8%; Score 141.8; DB 21; Length 1430;
Best Local Similarity 69.0%; Pred. No. 4.4e-23;
Matches 194; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
Db 960 GAAGCCCATTCACAGTGCACATCCCTGGTGGCGCAAGGTGTACGGCAAGGCTTCGCA 1019
867 GCAGAGGAGCACAGTGCATACGCGGGTGCAGAAAGTTTACGGAAATCTTCGCA 926
QY 1020 TGTGAAGCCCACTTGGCGTGGCACACTGGCGAGAGGCTTTCGTCGAACCTGGCTTTT 1079
Db 927 CCTCAAGGCGCACCTGAGAACTCACACAGGTGAGAGGCGCTTCGCTGCAAGTGCACGA 986
QY 1080 CTGCGGCAGAGGTTCACTGCTCTGACAGGTGGAGGCGCAGTGGCGACTCACACCG 1139
Db 987 CTGCAACAAGAAGTTGCGCGCTCCGACGAGCTGGCGGCGCACTACCGCACACACGCG 1046
QY 1140 GGAGAGAGAGTTCACCTTGCTGCTCTGTTCACAGCGCTTACCAAGAGCGACCTTGAG 1199
Db 1047 CGAGAGAGATTCAGTGCCTCATCTGGAGAGGCGCTTCATGCGCAGCGACCACTGAC 1106
QY 1200 CAACATCATCGCACCGGAGGAGCGGAGCGGCGGAGCGG 1240
Db 1107 CAAGCAGCGCGCGCGCGCACTTCACCGCGGAGTGC 1147

RESULT 12
AAAS0468
ID AAAS0468 standard; cDNA; 1471 BP.
XX
AC AAAS0468;
XX

05-DEC-2000 (first entry)
Mouse zinc finger transcription factor RFLAT-1 cDNA.
XX
KW RFLAT-1; RANTES factor of late activated T-lymphocyte;
KW transcription factor; zinc finger; mouse; inflammation;
KW antinflammatory; autoimmune disease; AIDS; neoplasia;
KW gene therapy; diagnosis; ss.
XX
OS Mus musculus.
XX
PN WO200052030-A1.
XX
PD 08-SEP-2000.
XX
PF 27-JAN-2000; 2000WO-US02300.
XX
PR 27-JAN-1999; 99US-0117576.
XX
PA (STRD) UNIV LELAND STANFORD JUNIOR.
XX
PI Song AM, Chen Y, Krensky AM;

XX WPI: 2000-587291/55.
DR New regulated upon activation, normal T cell expressed and secreted
XX (RANTES) factor of late activated T lymphocytes-1 proteins and nucleic
PT acids, for treating conditions associated with RANTES activity, e.g.
PT inflammation, AIDS
XX
XX Example; Page 47; 53pp; English.

This is the nucleotide sequence of a 1471 bp cDNA clone encoding
novel mouse RFLAT-1 (regulated upon activation, normal T cell
expressed and secreted (RANTES) factor of late activated T
lymphocytes-1). The cDNA was obtained by screening a murine
activated spleen cDNA library. At the protein level, mouse
RFLAT-1 shares about 90% identity with human RFLAT-1 (see AA95986).
RFLAT-1 is a novel zinc finger transcription factor that activates
RANTES gene expression to the A region of the RANTES gene promoter.
RFLAT-1 polypeptides and nucleic acids are used in diagnostic and
therapeutic agent screening, discovery, and preparation applications,
as well as in modulating RANTES activity, especially in the treatment
of conditions associated with RANTES activity, such as in the
suppression of RANTES gene expression, e.g. acute or chronic
inflammation, autoimmune diseases, and in conditions in which
enhancement of RANTES gene expression is needed, e.g. AIDS, or
neoplastic diseases. The RFLAT-1 gene, gene fragments, or the
encoded protein are useful in gene therapy to treat disorders
associated with gene defects or disease conditions in which
enhanced gene activity is desired. The gene may also be used in the
identification of RFLAT-1 gene homologues, as a source of novel
promoter elements, in the identification of RFLAT-1 expression
regulatory factors, to produce as probes and primers, in the
identification of expression patterns in a biological specimens,
and in the preparation of cell or transgenic animal models of
RFLAT-1 protein function.

SQ Sequence 1471 BP; 287 A; 549 C; 455 G; 180 T; 0 other;

Query Match 4.6%; Score 136.8; DB 21; Length 1471;
Best Local Similarity 60.5%; Pred. No. 6.6e-22;
Matches 242; Conservative 0; Mismatches 157; Indels 1; Gaps 1;

QY 960 GAAGCCCATTCACAGTGCACATCCCTGGTGGCGCAAGGTGTACGGCAAGGCTTCGCA 1019
Db 871 GCAGAGGAGCACAGTGCATACGCGGGTGCAGAAAGTTTACGGAAATCTTCGCA 930
QY 1020 TGTGAAGCCCACTTGGCGTGGCACACTGGCGAGAGGCTTTCGTCGAACCTGGCTTTT 1079
Db 931 CCTCAAGGCGCACCTGAGAACTCACACAGGTGAGAGGCGCTTCGCTGCAAGTGCACGA 990
QY 1080 CTGCGGCAGAGGTTCACTGCTCTGACAGCTGGAGCGCCACGTGCGCACTCACACCG 1139
Db 991 GTGCAACAAGAAGTTGCGACGCTCGGACGAGCTGGCGACGCACTATCGCACGACACGG 1050
QY 1140 GGAGAGAGTTCACCTTGCTGCTCTGTTCAGGCGCTTTCAGAGAGGCGACCACTTGAG 1199
Db 1051 CGAGAAGAAGTTCACTGCTGCCCATCTGTGAGAAGCGCTTCATCGGAGGAGCACTGAC 1110
QY 1200 CAACATCATCGCACCGGAGGAGCGGAGCGGCGGAGCGGCGGCGGCGGCGG 1258
Db 1111 GAAGCAGCAGCGCGCGCAGCGCAACTTCCACCGAGGATGCTGACGCGGCGGCGG 1170
QY 1259 AGCTGGGGAGGTCGCGAGCGTTCGGGGAAGAAAGACCAATCAGCGCGCCCGATCTTCCA 1318
Db 1171 CTCGAGGAGCGGCTCGCTCAGCGACTACAGCGGCTCCGATGCCAGCAGCGCCCATCAG 1230
QY 1319 CTTCCGCTGCACCCCGACAGAAAGCCACGGAGGAGCGCC 1358
Db 1231 CCGGCGCAGCTCACCTGAGCACCGCGCGCTGGACCGCGC 1270

RESULT 13
RAI60335

ID AAI60335 standard; cDNA: 3090 BP.
AC AAI60335;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 4324.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
QS Homo sapiens.
PA WO200153312-A1.
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 13-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
DR P-PSDB; AAM41179.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Claim 1; SEQ ID NO 4324; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
XX the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
XX
SQ Sequence 3090 BP; 814 A; 852 C; 752 G; 672 T; 0 other;

Query Match
Best Local Similarity 4.5%; Score 134.6; DB 22; Length 3090;
Matches 225; Conservative 63.4%; Pred. No. 2.8e-21;
Mismatches 124; Indels 6; Gaps 1;

QY 869 ATGCGGCGAGTGGGGCAGGGGCTTCTACCTGCGACTGCCCAACTGTCTCAGGAGCTAGAGC 928
DB 1884 AAGCCGGTCCGGAGGACCGGGGAGGAGCATGCACCTGCCCTCTGTATAGACAGTGAAG 1943

QY 929 GGCTCGGGCAGCAGCGGCTGGGCTGAGGAAAGAACGCCCATTCACAGCTGCCACATCCCTG 988
DB 1944 GAAGGGGCTCGGGGATCCTGCAAAAAGAACACATATT-----TGGCACATCCCAAG 1997
QY 989 GGTGCGCAAGGTGTACGGCAAGGCTTCGCATCTGAAAGCCACTTGGCTGGGCACTG 1048
DB 1998 GCTGTGGAAAGTGTATGGCAAGCCTCTCACCTCGGGGCACACTTGGCTGGCATACAG 2057
QY 1049 GCGAGAGCCCTTTCGTCTGCAACTGGCTTTTCTGCGGCAAGAGTTTCACTCTCTGACG 1108
DB 2058 GCGAGAGCCCTTTCGTCTGCAACTGGCTTTTCTGCGGCAAGAGTTTCACTCTCTGACG 2117
QY 1109 AGCTGGAGCGCCACGTGGCGCACTCACACCCGGGAGAAAGTTTCACTCTGCTGCTGT 1168
DB 2118 AGCTACAGAGGCAACAAAGCTACACACAGGTGAGAAGAAATTTGCTGCTGCTGCTGTC 2177
QY 1169 CCAAGCGCTTTACCGAAGCGACCACTTGAGCAACACATCAGCGCACCCACGGGGA 1223
DB 2178 CTAAGCGCTTCATGAGGAGTGACCACTGTCAAAACATATCAAGACCCACCGAGAA 2232

RESULT 14
AAI58549
ID AAI58549 standard; cDNA: 3289 BP.
XX
AC AAI58549;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 752.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
DR P-PSDB; AAM39393.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Claim 1; SEQ ID NO 752; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
XX the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
XX
SQ Sequence 3090 BP; 814 A; 852 C; 752 G; 672 T; 0 other;

of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localized neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: Immune system suppression, Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemia and C.N.S disorders.

CC Note: The sequence data for this patent did not form part of the printed specification.

XX Sequence 3289 BP; 867 A; 894 C; 786 G; 742 T; 0 other;

Query Match 4.5%; Score 134.6; DB 22; Length 3289;
Best Local Similarity 63.4%; Pred. No. 2.9e-21;
Matches 225; Conservative 0; Mismatches 124; Indels 6; Gaps 1;

QY 869 ATGCGGCGAGTGGGCGAGGCGCTTCTACTGCGACTGCCCACTGTCAGGAGCTAGAGC 928
DB 1981 AGCCCGTGGAGGACCGCGGGGAGCATGACCTGCCCTCTACTGTAAGACAGTGAAG 2040
929 GGTCTGGGCGACGAGCGGCTGGGCTGAGGAAGAGCCATTCACAGCTGCCACATCCCTG 988
DB 2041 GAAGGGCTCGGGGATCTCGCAAAAGAAACAGCATATT-----TGCCACATCCAG 2094
989 GGTGGGGAAGGTGACGCAAGGCTTGCATCTGAAGCCCACTGGCTGGCAGCTG 1048
DB 2095 GCTGTGGGAAGGTATGATGACAGAGCTCTCACTCGGGGACACATTCGCTGGCATAG 2154
QY 1049 GCGAGAGGCTTCTGCTCACTGCTTTCTGCGGAGAGGTTCACTGCTCTGAGC 1108
DB 2155 GCGAGAGGCTTCTGCTCACTGCTTTCTGCGGAGAGGTTCACTGCTCTGAGC 1168
QY 1109 AGCTGGAGCGCCAGCTGCGCTACACCGGGGAGGAAGTTCACTGCTGCTGTT 1168
DB 2215 AGCTACAGAGGCGCAACAGCTACACACAGGTGAGAGAAATTTGCTGCTGAGTGT 2274
QY 1169 CCAAGGCTTTACAGAGGAGGAGTGGAGCAATCTGAGCAATCTGAGGAGGGA 1223
DB 2275 CTAAGGCTTCTGAGGAGTGGAGGAGTGGAGGAGTGGAGGAGTGGAGGAGTGGAG 2329

RESULT 15

ABQ54824
ID ABQ54824 standard; cDNA; 1615 BP.

XX AC ABQ54824;

XX DT 22-AUG-2002 (first entry)

Human ovarian antigen HSQK12 cDNA, SEQ ID NO:704.

XX Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;
KW inflammatory condition; immune disorder; blood disorder;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disorder; urinary system disorder; drug screening;
KW gene therapy; chromosome mapping; forensic analysis;
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW antinflammatory; gynaecological; reproductive; gene; ss.

XX OS Homo sapiens.

XX PN WO200200677-A1.

XX PD 03-JAN-2002.

XX PF 07-JUN-2001; 2001WO-US18569.

PR 07-JUN-2000; 2000US-209467P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Birse CE, Rosen CA;

XX WPI: 2002-147878/19.

XX P-FSDB; ABP41747.

Isolated nucleic acid molecules encoding novel ovarian polypeptides, useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian cancer), immune disorders, cardiovascular disorders and neurological diseases -

PS Claim 1; SEQ ID No 704; 2922pp; English.

The invention relates to 2175 novel human ovarian antigens (ABP41054-ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also encompasses polypeptides 90% identical and polynucleotides 95% identical to the sequences of the invention. The invention additionally relates to recombinant vectors and host cells comprising human ovarian antigen polynucleotides, antibodies against human ovarian antigens, and the use of ovarian antigen polynucleotides and polypeptides in diagnosing, treating, prognosing or preventing various ovary and/or breast-related disorders. Such conditions include ovarian cancer and breast cancer, and metastatic tumours of ovarian or breast origin, reproductive system disorders (e.g., infertility, disorders of pregnancy, anovulation, polycystic ovary syndrome, ovarian cysts, and dysmenorrhea), endocrine disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and vaginitis), immune disorders (e.g., congenital and acquired immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus), blood-related disorders (e.g., anaemia), cardiovascular disorders, respiratory disorders, neurological disorders, gastrointestinal disorders and urinary system disorders. Ovarian antigen polypeptides and polynucleotides may also be used in screening for compounds which modulate ovarian antigen expression or activity. The polynucleotides may further be used for gene therapy, chromosome mapping, in the identification of individuals and in forensic analysis, and the polypeptides may be used as food additives or to prepare antibodies useful in disease diagnosis, drug targeting and phenotyping. The present sequence represents cDNA encoding a human ovarian antigen of the invention.

CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1615 BP; 383 A; 478 C; 356 G; 394 T; 4 other;

Query Match 4.4%; Score 130.8; DB 24; Length 1615;

Best Local Similarity 66.9%; Pred. No. 1.7e-20;

Matches 186; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 945 GGCTGGCTGAGGAAGAGCCCATTCACAGCTGCCACATCCCTGGTGGCGGAAGTGA 1004

DB 148 GTCTGGAGAGGAGGAGCAAGAACAGCAGTGTGCCACATCCCGACTGTGGCAAGACGTT 207

QY 1005 CGCGAAGGCTTCGCTGTAAGCCCACTTGGCGTGGCAGCTGGCGAGAGGCTTTCGT 1064

DB 208 CGCTAAGACGCTCTTGTCTGCTGGTGGCCATGTGCGCTGCACACTGGGAGCGGCTTCT 267

QY 1065 CTGCAACTGGCTTTCTTCTGGGCAAGAGTTCAGTCTGCTGAGAGCTGAGGCGGCACT 1124

DB 268 CTGCAACTGGTTCCTTCTGTGGGAAGAGTTTACACAGGAGTACAGAGCTCAACGCGATC 327

QY 1125 GCGCACTCACACCGGAGAGAGTTCAGTTCGCTGCTCTCTCAAGCGCTTTACCA 1184

DB 328 TCGCACCCACACAGGGGACAAACGCTTCAGTGGCGCCAGTGTACAGAGCGCTTCATGAG 387

QY 1185 AAGCGACCACTTGTAGCAAAACATCAGCGCACCCACCGGG 1222

DB 388 GAGTGACCACTCACCACGATTACAGAGACCCACCTGG 425

Search completed: February 21, 2003, 22:19:18
JOB time : 440 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: February 21, 2003, 16:47:28 ; Search time 2676 Seconds
(without alignments)
17914.290 Million cell updates/sec

Title: US-09-734-329-1
Perfect score: 2960
Sequence: 1 attctccatttccctccc.....aaaaaaaaaaaaaaaaaaaaa 2960

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
rched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estinu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_esti:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	798.4	27.0	970	9	AV227235
2	672.2	22.7	707	10	BB028252
3	624.8	21.1	637	10	BB048324
4	607.6	20.5	615	10	BB622399
5	591	20.0	652	10	AV337525
6	502	17.0	515	9	AA034571

7	455.8	15.4	643	10	BB622895	
8	447	15.1	455	10	AW743495	
9	437.6	14.8	661	14	BQ205022	
c	10	428.2	14.5	681	13	BM389110
11	424.8	14.4	430	10	BB675901	
12	421.2	14.2	428	10	BB674942	
13	401.8	13.6	435	10	BB689394	
14	394.6	13.3	670	10	BB618195	
c	15	379	12.8	594	13	B1303918
c	16	359.4	12.1	562	13	BM382845
17	321.6	10.9	520	12	BF564272	
c	18	310.2	10.5	465	12	BF564310
19	287.4	9.7	322	10	BB363167	
20	272.2	9.2	307	10	BB132968	
21	266.8	9.0	290	10	BB470471	
22	265.2	9.0	339	10	BB131724	
23	261.6	8.8	296	10	BB527673	
24	257.4	8.7	300	10	BB175776	
25	253.8	8.6	290	10	BB368597	
26	251.2	8.5	280	10	BB360632	
27	249.2	8.4	289	10	BB375475	
28	248.2	8.4	298	10	BB526105	
29	247.2	8.4	678	14	BQ183701	
c	30	245	8.3	301	10	BB133795
31	242	8.2	282	10	BB521809	
32	241.2	8.1	407	12	BF407895	
33	238.6	8.1	315	9	AV235790	
34	234.4	7.9	285	9	AV204031	
35	226.2	7.6	381	10	AW534398	
36	226	7.6	263	10	BB354942	
37	224	7.6	260	10	BB523022	
38	223.4	7.5	319	10	BB374512	
39	215.2	7.3	298	10	BB525281	
40	209	7.1	356	10	AW532290	
41	205.8	7.0	252	9	AV245678	
42	205.4	6.9	242	9	AV243081	
43	205	6.9	230	10	BB035617	
c	44	204	6.9	1798	11	BC023224
c	45	198.8	6.7	647	9	AL662349

ALIGNMENTS

RESULT 1
AV227235
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AV227235 970 bp mRNA linear EST 14-NOV-2001
musculus cDNA clone 412402c11 3', mRNA sequence.
AV227235.2 GI:16385425
house mouse.
Mus musculus.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 970)
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
On Nov 1, 1999 this sequence version replaced gi:6178902.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222

Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.
Unpublished (2001)
On Jun 7, 2000 this sequence version replaced gi:8323892.

TITLE JOURNAL COMMENT

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@sc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanuki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Yamanaka, I., Kiyosawa, H., Kondo, S., Saito, T., Shinagawa, A., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K., Arakawa, T., Ishii, Y. and Hayashizaki, Y.
Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct. Funct. Genomics 2 pre, L72-L86 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

FEATURES source

Location/Qualifiers
1. 707
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="RIKEN full-length enriched, 6 days neonate head"
/sex="mixed"
/tissue_type="head"
/dev_stage="6 days neonate"
/lab_host="B610b"
/note="Site_1: Salt; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGATCCAGAGCTCTTTTATTTTNN 3'], cDNA was prepared by using trehalase thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCGAGTAAATTAATCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC 1. Cloning sites, 5' end: Salt; 3' end: BamHI"

BASE COUNT 180 a 164 c 162 g 201 t
ORIGIN

Query Match 22.7%; Score 672.2; DB 10; Length 707;
Best Local Similarity 97.9%; Pred. No. 4e-95;
Matches 692; Conservative 0; Mismatches 13; Indels 2; Gaps 1;

```
QY 2054 ATAATCTATTATTAAGTATGATGACGGGTGAGTACAGTACAGTGGAGAGGAAA- 2112
Db 1 AAAATCTATTATTAAGTATTAATGATGACGGGTGAGTAAAGTACAGTGGAGAGGAAA 60

QY 2113 -GGATTTCTCCCGCCCCCAAGGAATCTAGTCAATGATCTCTGTATAGACAAATGA 2171
Db 61 GGGATTTCTCCCGCCCCCAAGGAATCTAGTCAATGATCTCTGTATAGACAAATGA 120

QY 2172 TAGTGGAGACCTTGCTCGTAGATTCTATCCTCGAGGTCTCCGAGAGTCTCTTTTTCAGT 2231
Db 121 TAGTGGAGACCTTGCTCGTAGATTCTATCCTCGAGGTCTCCGAGAGTCTCTTTTTCAGT 180

QY 2232 TGAGTTTGGGTGTTGTCGGGCTCTTTTAGAGTTTCTGTGGGTGTCCTCTCTGTAGGCAGT 2291
Db 181 TGAGTTTGGGTGTTGTCGGGCTCTTTTAAAGTTTCTGTGGGTGTCCTCTCTGTAGGCAGT 240

QY 2292 CACTAAGATCCCAGCCCCCAGCAGAAAGCTGTGAAACTTCAAGTCTCTATGGCGGGAGG 2351
Db 241 CACTAAGATCCCAGCCCCCAGCAGAAAGCTGTGAAACTTCAAGTCTCTATGGCGGGAGG 300

QY 2352 ACTGGAATGTACCCAGCTCTCGACCCGACTGCAGATCAGGTCTCCCTCATCTCCTC 2411
Db 301 ACTGGAATGTACCCAGCTCTCGACCCGACTGCAGATCAGGTCTCCCTCATCTCCTC 360

QY 2412 TTCTCATACCTGTGACCTCACCAGGTTATCCCTTGTGTCATGTTTACAGAGAGTTG 2471
Db 361 TTCTCATACCTGTGACCTCACCAGGTTATCCCTTGTGTCATGTTTACAGAGAGTTG 420

QY 2472 CAGCTGCCATCTTAAGAGTGTCTTTGGGGAGAGCCACCTACAGAGAGATTTGGTT 2531
Db 421 CAGCTGCCATCTTAAGAGTGTCTTTGGGGAGAGCCACCTACAGAGAGATTTGGTT 480

QY 2532 TGGAGTGCCCTCCCTCAAAAAGTAGTGGCAAGGCTTTCTCTGGGATCAAAATTCAAA 2591
Db 481 TGGAGTGCCCTCCCTCAAAAAGTAGTGGGCAAGGCTTTCTCTGGGATCAAAATTCAAA 540

QY 2592 TAAATCAAGTATTATTGAATGCTTAATATGTGCAAGGCTGTGTCCTAGAGCCAGAG 2651
Db 541 TAAATCAAGTATTATTGAATGCTTAATATGTGCAAGGCTGTGTCCTAGAGCCAGAG 600

QY 2652 AAGAATTTATACAGACAGAGAGTCCCTTAACATAACATCCACAGGCCCAATCTAGG 2711
Db 601 AAGAATTTATACAGACAGAGAGTCCCTTAACATAACATCCACAGGCCCAATCTAGG 660

QY 2712 AGGTTTCACCTCCATTCAGTGACTTTTAAAGCCGCTTTGTGCGCTTTG 2758
Db 661 AGGTTTCACCTCCATTCAGTGACTTTTAAAGCCGCTTTGTGCGCTTTG 707
```

RESULT 3
BB048324

LOCUS
BB048324

DEFINITION
musculus cDNA clone 6430578P22 3', mRNA sequence.

ACCESSION
BB048324

VERSION
BB048324.2

KEYWORDS
EST.

SOURCE
house mouse.

ORGANISM
Mus musculus

REFERENCE
1 (bases 1 to 637)

AUTHORS
Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki

637 bp mRNA linear EST 18-OCT-2001
BB048324 RIKEN full-length enriched, adult male olfactory bulb Mus

musculus cDNA clone 6430578P22 3', mRNA sequence.

EST.

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 637)

Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki


```
source 1. 515
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_image="IMAGE:46845"
/clone_lib="Soares mouse embryo NbME13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/notes="vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTACCAACTGAGTGGAGGCGCGGAAATTTTTTTTTTTTTTTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos (total RNA provided by Minoru Ko, Wayne
State Univ., from 2); double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."
BASE COUNT 124 a 122 c 122 g 147 t
ORIGIN

Query Match 17.08; Score 502; DB 9; Length 515;
Best Local Similarity 100.0%; Pred. No. 1.3e-68;
Matches 502; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2002 GTACTTAACCCCTATGCCCCAGGAGTAAGTAATAGTATAGTAATAATAAATCT 2061
|||||
DB 14 GTACTTAACCCCTATGCCCCAGGAGTAAGTAATAGTATAGTAATAATAAATCT 73
|||||

QY 2062 ATTTATCTAAGTTATGATGACGGTTCAGTACAGTCTGAGAGGAGAAAGGATCTC 2121
|||||
DB 74 ATTTATCTAAGTTATGATGACGGTTCAGTACAGTCTGAGAGGAGAAAGGATCTC 133
|||||

QY 2122 CCGCGCCCAAGGAATCTAGTCAAAATGATCTCTGTATAGACAAATGATGAGAC 2181
|||||
DB 134 CCGCGCCCAAGGAATCTAGTCAAAATGATCTCTGTATAGACAAATGATGAGAC 193
|||||

QY 2182 CTTCGCTGATAGTTCTATCTCCGAGGTCCTCCGAGAGTTCTTTTTCAGTTGAGTTGG 2241
|||||
DB 194 CTTCGCTGATAGTTCTATCTCCGAGGTCCTCCGAGAGTTCTTTTTCAGTTGAGTTGG 253
|||||

2242 GTTGTGGGCTCTTTAGAGTTCTGTGGGTGCTCTCTCTTAGGAGTCACATAAGTC 2301
|||||
DB 254 GTTGTGGGCTCTTTAGAGTTCTGTGGGTGCTCTCTCTTAGGAGTCACATAAGTC 313
|||||

QY 2302 CCCAGCCCCAGCAGAAAGCTGTGAACTTCAAGTCTATGCGCGGGAGGACTGGAATGT 2361
|||||
DB 314 CCCAGCCCCAGCAGAAAGCTGTGAACTTCAAGTCTATGCGCGGGAGGACTGGAATGT 373
|||||

QY 2362 ACCCGAGTCTCTCGACCCGAGTACAGATCAGTTCTCCCTGATCTCTCTTCATACC 2421
|||||
DB 374 ACCCGAGTCTCTCGACCCGAGTACAGATCAGTTCTCCCTGATCTCTCTTCATACC 433
|||||

QY 2422 CTGTGACCTACACAGTTATCCCTTGTGTCATGTTGTTACAGAGAGCTTCGAGTGCAT 2481
|||||
DB 434 CTGTGACCTACACAGTTATCCCTTGTGTCATGTTGTTACAGAGAGCTTCGAGTGCAT 493
|||||

QY 2482 CTAAACGCTCTCTTTGGGGA 2503
|||||
DB 494 CTAAACGCTCTCTTTGGGGA 515
|||||

RESULT 7
LOCUS BB622895 643 bp mRNA linear EST 26-OCT-2001
DEFINITION BB622895 RIKEN full-length enriched, adult male olfactory brain Mus
musculus cDNA clone 6430578P22 5', mRNA sequence.
ACCESSION BB622895
VERSION BB622895.1 GI:16461780
```

KEYWORDS
SOURCE
ORGANISM

EST.
house mouse.
Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS

1 (bases 1 to 643)
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)

TITLE
JOURNAL
COMMENT

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10): 1617-1630 (2000)
wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11): 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2): 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-length Mouse cDNAs Compared with
Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.

FEATURES
source

Location/Qualifiers
1. 643
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="RIKEN full-length enriched, adult male
olfactory brain"
/sex="male"
/tissue_type="olfactory brain"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site 1: SalI; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGCGCGCAACTCGAGTTTTTTTTTTTTT 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 100.0. Second
strand cDNA was prepared with the primer adaptor of

JOURNAL
MEDLINE
COMMENT

Genome Res. 6 (9), 791-806 (1996)
9704477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dt track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dt track served to verify it as a clone from the
normalized cartilaginous tumor library cDNA Library Preparation:
M.B. Soares Lab Clone Distribution: clones will be available
through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes

FEATURES
source

Location/Qualifiers
1. .661
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-D21-cn1-a-06-0-UI"
/clone_lib="UI-R-D21"
/tissue_type="Chondrosarcoma"
/dev_stage="37 days"
/lab_host="DH10B (Life Technologies)"
/note="Organ: Spine; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI;
UI-R-D21 is a normalized cDNA library containing the
following tissue(s): Swam Rat Chondrosarcoma. The library
was constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dt primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT73D-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
CATCTGTGTA. The Rat cartilaginous tumor tissue was
provided by Dr. Jeff Stevens at the University of Iowa.
TAG_L1B=UI-R-D21
TAG_TISSUE=cartilaginous tumor
TAG_SEQ=CATTCTGTGA"
BASE COUNT 160 a 145 c 151 g 205 t
ORIGIN

Query Match 14.8%; Score 437.6; DB 14; Length 661;
Best Local Similarity 89.6%; Pred. No. 1.2e-58;
Matches 577; Conservative 0; Mismatches 44; Indels 23; Gaps 9;
QY 2129 CAAGGAATCTAGTCAAAATGCATCTCTGT----ATAGACAATGATAGTGGAGACCTT 2184
DB 22 CTAAGGGTATTAGACAAAATGCATCTGTGTATAGATAGACAAAATGATAGTGGAGACCT- 80
QY 2185 GCTCTAGATTCTATCTCTCGAGGCTCCGAGAGTCTCTTTTCAGTTGAGTTGGGTT 2244
DB 81 --TCGTAGATTCTATCATCGAGTCTCT--GAGTTCTTTTTCAGCGAGGATTTGGGTT 136
QY 2245 GTTCGGCTCTTTTATAGATTTCTGTGGGTCTCTCTCTGTAGGCACTCAATAGATCCCC 2304
DB 137 GTTAGGCTCTTTT--AGTTATGTGGGTCTCTCTCTGTAGGCACTCAATAGATCTCA 194
QY 2305 AGCCCCAGCCAGAGCTGTGAATCTCAAGTCTATATCGCGGGAGGAGTGAATGTACC 2364
DB 195 AGCCCCAGCCAG-AGCTGTGAAATCTCAAGTCTATATGAGGGAGGAGTGAATGTACC 253
QY 2365 CCAGTCTCTCGACCCGACTCGAGATCATGAGTTCTCTCCCTCATCTCTTCATCATCCGTG 2424
DB 254 CCAGTCTCTTGACC-----ACAGGAGAGGTTCTTCAACTGACCTCTCTCATACCCGTG 308

QY 2425 TGACCTCACCAGTTATCCCTTGTCATGTTTACAGAGCTTGAGCTGCCATCTT 2484
DB 309 TGACCTCACCAGTTATCCCTTGTCATGTTTATGAGAGCTTGAGCTGCCATCTT 368
QY 2485 AAACGTGCTCTTTGGGGAGAGCCACCATACAGAGAGATTTGGTTGGAGGTGCCCT 2544
DB 369 AAACGAGTTCTTT---GGAGAGCCCATCTAACAGGAGGATTTGGTTGGAGGTGCCCT 425
QY 2545 CCTGAAAAGTAGTGGGCAAGGCTTCTCTGGCATCAAAATCAAAATAAACAAGTATT 2604
DB 426 CCTGAAAAGTAGTGGGCAAGGCTTCTCTGGCATCAAAATCAAAATAAACAAGTATT 485
QY 2605 TATTGAATGCTTAATATGTCAAGGCTTGCTGCTTAGAGCCACGAGAAAGATTTATAA 2664
DB 486 TATTGAGTGCTTACTATGTGCAAGGCTTGCTGCTTAGAGCCATGAGAAAGATTTATAA 545
QY 2665 CAGACAGAAGTCCCTTAAAC--TAAACATCCACAGGCCCTCAATCTAGGAGGTTTCACTC 2722
DB 546 CAGGACAGAGGTCCCTTAAACATATAAACAATCCACAGTCCCTCAATCTAGGAGGTTTCACTC 605
QY 2723 CATTCCAGTGAATTTTAAAGCGCTTTTGTGCTTTTGAATGCTT 2766
DB 606 CACTCCAGTGAATTTTAAAGCGCTTGTGCTTTTGAATGCTT 648
RESULT 10
BM389110/c
LOCUS BM389110 681 bp mRNA linear EST 17-JAN-2002
DEFINITION UI-R-D20-cko-c-07-0-UI.sl UI-R-D20 Rattus norvegicus cDNA clone
UI-R-D20-cko-c-07-0-UI 3', mRNA sequence.
ACCESSION BM389110
VERSION BM389110.1 GI:18189163
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 681)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 9704477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dt track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dt track served to verify it as a clone from the
non-normalized cartilaginous tumor library cDNA Library
Preparation: M.B. Soares Lab Clone Distribution: clones will be
available through Research Genetics (www.resgen.com) The following
repetitive elements were found in this cDNA sequence: 44-96,
>RMER17ChLTR_5/6_bp_duplication
Seq primer: M13 Forward
POLYA=Yes
Location/Qualifiers
1. .661
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-D20-cko-c-07-0-UI"
/clone_lib="UI-R-D20"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"

Email: genome-res@gsc.riken.go.jp,
URL: <http://genome.gsc.riken.go.jp/>

Research Genetics (www.reagen.com) The following repetitive
elements were found in this cDNA sequence: 44-96,
>RMR17C#LTR_5/6_bp_duplication
seq primer: M13 Forward
POLYA=Yes.

FEATURES

Location/Qualifiers
1..594
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-DR0-cjb-1-10-0-UI"
/clone_lib="UI-R-DR0"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-DR0
library is a non-normalized Rat Osteoblast library
constructed in pT377 PAC vector according to the procedure
described by Bonaldo, Lennon & Soares (Genome Research
Genome 6: 791-806, 1996). The oligonucleotide used to
prime first strand synthesis contained the sequence tag
CCAGG between the Not I cloning site and dt18 stretch.
TAG LIB=UI-R-DR0
TAG TISSUE=osteoblast
TAG_SEQ=AAAGATACAA"

BASE COUNT 177 a 114 c 146 g 155 t 2 others
ORIGIN

Query Match 12.8%; Score 379; DB 13; Length 594;
Best Local Similarity 89.9%; Pred No. 1.6e-49;
Matches 463; Conservative 0; Mismatches 42; Indels 10; Gaps 5;

QY 2394 GTCTCTCCCTGATCCTTCATACCCCTGTGACCTACACAGGTATCCCTTGTGCTC 2453
DB 594 GTCTTCAACTGACCTCTCTCATACCCCTGTGACCTATCAGGCTATTCCTCCTGTC 535
QY 2454 ATGGTTACAGAGCTTGCGAGTCCCATCTTAACGTGCTCTTTGGGGGAGAGCCACCT 2513
DB 534 ATGGTTATGGAGAGCTTGCGAGTCCCATCTTAACGTGCTCTTTGGGGGAGAGCCACCT 478
QY 2514 AACAGGAGGATTTGGTTGGAGGTGCCCTCTCTGAAAGTAGGTGGGCAAGGCTTTC 2573
DB 477 AACAGGAGGATTTGGTTGGAGGTGCCCTCTCTGAAAGTAGGTGGGCAAGGCTTTC 418
QY 2574 TCTGGGATCAATTCAAATTAATCAAGTATTTATTGAATGCTTAATATGTGCAAGGCTG 2633
DB 417 TCTGGGATCAATTCAAATTAATCAAGTATTTATTGAATGCTTAATATGTGCAAGGCTG 358
QY 2634 GTCCCTAGAGCCACAGAGAAAGATTTATACAGGACAGAGTCCCTAAACT--AAACAT 2691
DB 357 GTCCCTAGAGCCACAGAGAAAGATTTATACAGGACAGAGTCCCTAAACTATAACAT 298
QY 2692 CCACAGGCCCAATCTAGAGGTTCCTACCTCCATTCAGTCTCTTTTAAAGCCGCTTGT 2751
DB 297 CCACAGTCCCAATCTAGAGGTTCCTACCTCCATTCAGTCTCTTTTAAAGCCGCTT--G 240
QY 2752 GCCTTTGAATGCTTTCCTGAGATTTTGGATCTTCTGTCCTCCCTGCTCTCTCT 2811
DB 239 TGCCTTTGAATGCTTTCCTGAACTTTTGGATCTTCTGTCCTCCCTGCTCTCTCT 180
QY 2812 AGCCCTCAAGATTAAGGGTAAAGCCA--TGGAGTCTGGGAGAGC--ATAAGCTGTTGAC 2868
DB 179 AGCCCTCAAGATTAAGGGTAAAGCCAGTGGAGTCTGGGAGGGGCATATAACACCGTTGAA 120
QY 2869 GGGATCGTCCCTTGTGGATCTTTCTTTTCTTTT 2903
DB 119 GGGATCATCCCTTGTGGATCTTTCTCTCTCT 85

Search completed: February 21, 2003, 23:04:10
Job time : 2703 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 19, 2003, 13:32:31 ; Search time 40 Seconds

(without alignments)
1425.782 Million cell updates/sec

Title: US-09-734-329-2

Perfect score: 428

Sequence: 1 MASSLLLEERHYGSSPLAML.....PAPPEKAHGSPQSNLLEI 428

Scoring table:

Gapop 60.0 , Gapext 60.0

Arched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	7.7	666	22	ABBS8019
2	25	5.8	398	23	AA049116
3	25	5.8	398	23	AA049117
4	20	4.7	554	22	ABG15134
5	14	3.3	968	22	ABBS1691
6	12	2.8	66	22	AA073418
7	12	2.8	66	23	ABG43277
8	12	2.8	90	23	ABP48203
9	12	2.8	94	21	AA014293
10	12	2.8	94	23	ABJ03953

11	12	2.8	94	23	ABP48202
12	12	2.8	100	21	AA014294
13	12	2.8	100	23	ABJ03954
14	12	2.8	135	23	ABP41747
15	12	2.8	168	17	AA095975
16	12	2.8	241	17	AA096015
17	12	2.8	273	17	AA096011
18	12	2.8	297	18	AA089955
19	12	2.8	303	18	AA089956
20	12	2.8	612	23	ABG62110
21	12	2.8	785	22	AA039393
22	12	2.8	794	22	AA041179
23	11	2.6	97	22	AA062171
24	11	2.6	97	22	AA062172
25	11	2.6	99	21	AA070699
26	11	2.6	99	21	AA070700
27	11	2.6	99	22	AA08712
28	11	2.6	99	22	AA08713
29	11	2.6	99	23	AA021124
30	11	2.6	99	23	AA071131
31	11	2.6	99	23	AA071135
32	11	2.6	196	21	AA07701
33	11	2.6	196	22	AA08714
34	11	2.6	196	23	AA021123
35	11	2.6	196	23	AA071136
36	11	2.6	644	22	AB070436
37	10	2.3	397	22	ABG62783
38	10	2.3	1267	22	ABG61050
39	9	2.1	365	21	AA093317
40	9	2.1	409	22	AA075181
41	9	2.1	413	22	AB031159
42	9	2.1	413	22	AB036354
43	9	2.1	413	22	AB021712
44	9	2.1	413	22	AA057121
45	9	2.1	413	22	AA069513

ALIGNMENTS

RESULT 1
ABB58019
ID ABB58019 standard; Protein: 666 AA.
AC ABB58019;
XX
XX 26-MAR-2002 (first entry)
DE Drosophila melanogaster polypeptide SEQ ID NO 849.
KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
PF 23-MAR-2001; 2001WO-US09231.
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP. NY.
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
DR N-PSDB; ABL02122.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

PS Disclosure; SEQ ID NO 849; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from prosochilla. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (AB57737-AB972072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 666 AA;

Query Match 7.7%; Score 33; DB 22; Length 666;
Best Local Similarity 100.0%; Pred. No. 9.3e-23;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

306 SHLKAHLRHTGCTGPFVNCWLFPCGRFTRSDLEL 338

351 SHLKAHLRHTGCTGPFVNCWLFPCGRFTRSDLEL 383

RESULT 2

AA049116
ID AA049116 standard; Protein; 398 AA.

XX AC AA049116;

XX DT 20-MAY-2002 (first entry)

XX DE Mouse Spl family transcription factor m285.

XX KW Mouse; murine; m285; transcription factor; Spl family; zinc finger;
KW GC-box binding; 285 protein; promoter activity regulation;
KW transcriptional regulation; drug screening; drug development;
KW gene therapy.

XX OS Mus musculus.

Key	Location/Qualifiers
FT Region	164..173
FT	/label= Pro-rich_region
FT Region	226..235
FT	/label= Ala-rich_region
FT Domain	298..320
FT	/label= zinc_finger_1
FT Domain	328..350
FT	/label= zinc_finger_2
FT Domain	358..378
FT	/label= zinc_finger_3

XX WO200198492-A1.

XX PD 27-DEC-2001.

XX PF 14-JUN-2001; 2001WO-JP05066.

XX PR 20-JUN-2000; 2000JP-0189762.

XX PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

XX PI Senoo C, Numata M;

XX DR WPI; 2002-114585/15.

XX DR N-PSDB; ABA96948.

XX PT Spl family transcription factor having zinc finger domain for use as a
target for drug development

XX

PS Claim 1; Fig 1; 96pp; Japanese.

XX This sequence represents a murine Spl family transcription factor,
CC m285, and its human homologue, h285 (AA049117) contain 3
CC zinc finger domains and are present in the nucleus, where they bind GC
CC boxes and regulate transcriptional activity. The invention also
CC encompasses fragments of the 285 proteins and nucleic acids encoding
CC them, expression vectors and host cells comprising 285 protein-encoding
CC nucleic acids, the recombinant production of 285 proteins, antibodies
CC against 285 proteins, a method of screening for compounds which bind to
CC 285 proteins, and a method for the regulation of the transcription
CC activity of promoters, especially of the CMV promoter, using the 285
CC protein. The 285 proteins can be used target molecules for drug
CC development and for gene therapy.

SQ Sequence 398 AA;

Query Match 5.8%; Score 25; DB 23; Length 398;
Best Local Similarity 100.0%; Pred. No. 2.8e-15;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 306 SHLKAHLRHTGCTGPFVNCWLFPCGK 330

DB 311 SHLKAHLRHTGCTGPFVNCWLFPCGK 335

RESULT 3

AA049117
ID AA049117 standard; Protein; 398 AA.

XX AC AA049117;

XX DT 20-MAY-2002 (first entry)

XX DE Human Spl family transcription factor h285.

XX KW Human; h285; transcription factor; Spl family; zinc finger;
KW GC-box binding; 285 protein; promoter activity regulation;
KW transcriptional regulation; drug screening; drug development;
KW gene therapy.

XX OS Homo sapiens.

Key	Location/Qualifiers
FT Region	164..173
FT	/label= Pro-rich_region
FT Region	226..235
FT	/label= Ala-rich_region
FT Domain	298..320
FT	/label= zinc_finger_1
FT Domain	328..350
FT	/label= zinc_finger_2
FT Domain	358..378
FT	/label= zinc_finger_3

XX WO200198492-A1.

XX PD 27-DEC-2001.

XX PF 14-JUN-2001; 2001WO-JP05066.

XX PR 20-JUN-2000; 2000JP-0189762.

XX PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

XX PI Senoo C, Numata M;

XX DR WPI; 2002-114585/15.

XX DR N-PSDB; ABA96949.

XX PT Spl family transcription factor having zinc finger domain for use as a
target for drug development

XX PS Claim 1; Fig 6; 96pp; Japanese.

XX CC This sequence represents a human Sp1 family transcription factor,

XX CC h285, h285, and its murine homologue, m285 (AAM49116) contain 3

XX CC zinc finger domains and are present in the nucleus, where they bind GC

XX CC boxes and regulate transcriptional activity. The invention also

XX CC encompasses fragments of the 285 proteins and nucleic acids encoding

XX CC them, expression vectors and host cells comprising 285 protein-encoding

XX CC nucleic acids, the recombinant production of 285 proteins, antibodies

XX CC against 285 proteins, a method of screening for compounds which bind to

XX CC 285 proteins, and a method for the regulation of the transcription

XX CC activity of promoters, especially of the CMV promoter, using the 285

XX CC protein. The 285 proteins can be used target molecules for drug

XX CC development and for gene therapy.

XX SQ Sequence 398 AA;

Query Match 5.8%; Score 25; DB 23; Length 398;

Best Local Similarity 100.0%; Pred. No. 2.8e-15;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 306 SHLKAHLRWHWTGPRFVCNWLFGCK 330

DB 311 SHLKAHLRWHWTGPRFVCNWLFGCK 335

RESULT 4

ABB15134

ID ABB15134 standard; Protein: 554 AA.

XX AC ABB15134;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #15125.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

AA 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI: 2001-639362/73.

DR N-PSDB; AAS79321.

XX PT New isolated polynucleotide and encoded polypeptides, useful in

XX PT diagnostics, forensics, gene mapping, identification of mutations

XX PT responsible for genetic disorders or other traits and to assess

XX PT biodiversity.

XX PS Claim 20; SEQ ID No 45493; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and

XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

XX CC and gene mapping, and in recombinant production of (II). The

XX CC polynucleotides are also used in diagnostics as expressed sequence tags

XX CC for identifying expressed genes. (I) is useful in gene therapy techniques

XX CC to restore normal activity of (II) or to treat disease states involving

XX CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG30377 represent novel human

CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 554 AA;

Query Match 4.7%; Score 20; DB 22; Length 554;

Best Local Similarity 100.0%; Pred. No. 2.4e-10;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 319 RPFVCNWLFGCKRFRSDEL 338

DB 358 RPFVCNWLFGCKRFRSDEL 377

RESULT 5

ABB61691

ID ABB61691 standard; Protein: 968 AA.

XX AC ABB61691;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 11865.

XX KW Drosophila; developmental biology; cell signalling; insecticide;

XX KW pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI: 2001-656860/75.

DR N-PSDB; ABL05794.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more

XX PT genes from Drosophila and for elucidating cell signalling and cell-cell

XX PT interactions.

XX PS Disclosure; SEQ ID NO 11865; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent

XX CC capable of detecting 1000 or more genes from Drosophila. The invention is

XX CC useful in developmental biology and in elucidating cell signalling and

XX CC cell-cell interactions in higher eukaryotes for the development of

XX CC insecticides, therapeutics and pharmaceutical drugs. The invention

XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

XX CC sequences (ABL01840-ABL16175) and the encoded proteins

XX CC (ABBS7737-ABBS72072).

XX CC The sequence data for this patent did not form part of the printed

XX CC specification, but was obtained in electronic format directly from WIPO

XX CC at ftp.wipo.int/pub/published_pct_sequences.


```

XX SQ Sequence 968 AA;
Query Match 3.3%; Score 14; DB 22; Length 968;
Best Local Similarity 100.0%; Pred. No. 0.00024;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 310 AHLRWHTGERPFVC 323
Db 759 AHLRWHTGERPFVC 772
|||||

RESULT 6
ID AAM73418 standard; Protein; 66 AA.
XX AC AAM73418;
XX DT 06-NOV-2001 (first entry)
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 33724.
XX Human; bone marrow expressed exon; gene expression analysis; probe;
microarray; cancer; leukaemia; lymphoma; myeloma.
XX Homo sapiens.
XX OS
XX PN WO200157276-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00668.
XX PA
XX PI 04-FEB-2000; 2000US-0180312.
XX PI 26-MAY-2000; 2000US-0207456.
XX PI 30-JUN-2000; 2000US-0608408.
XX PI 03-AUG-2000; 2000US-0632366.
XX PI 21-SEP-2000; 2000US-0234687.
XX PI 27-SEP-2000; 2000US-0236359.
XX PI 04-OCT-2000; 2000US-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow.
XX Example 4; SEQ ID NO: 33724; 658pp + Sequence Listing; English.

The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention.

XX SQ Sequence 66 AA;
Query Match 2.8%; Score 12; DB 22; Length 66;
Best Local Similarity 100.0%; Pred. No. 0.0017;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 310 AHLRWHTGERPF 321
Db 30 AHLRWHTGERPF 41
|||||

RESULT 7
ID ABG43277 standard; Peptide; 66 AA.
XX AC ABG43277;
XX DT 19-AUG-2002 (first entry)
XX DE Human peptide encoded by genome-derived single exon probe SEQ ID 32942.
XX Human; single exon probe; asthma; lung cancer; COPD; ILD;
XX chronic obstructive pulmonary disease; interstitial lung disease;
XX familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
XX pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
XX primary ciliary dyskinesia; pulmonary hypertension;
XX hyaline membrane disease.
XX Homo sapiens.
XX OS
XX PN WO200186003-A2.
XX PD 15-NOV-2001.
XX PF 30-JAN-2001; 2001WO-US00665.
XX PA
XX PI 04-FEB-2000; 2000US-180312P.
XX PI 26-MAY-2000; 2000US-207456P.
XX PI 30-JUN-2000; 2000US-0608408.
XX PI 03-AUG-2000; 2000US-0632366.
XX PI 21-SEP-2000; 2000US-234687P.
XX PI 27-SEP-2000; 2000US-236359P.
XX PI 04-OCT-2000; 2000US-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2002-114183/15.
XX Spatially-addressable set of single exon nucleic acid probes, used to
PT measure gene expression in human lung samples.
XX Claim 27; SEQ ID NO 32942; 634pp; English.

The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements or the 12387 open reading frames derived from the 12614
CC probes. Also included are a microarray comprising the novel set of
CC probes; the novel set of probes which hybridize at high stringency to a
CC nucleic acid expressed in the human lung; measuring gene expression in a
CC sample derived from human lung, comprising (a) contacting the array with
CC a collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of
CC the array; identifying exons in a eukaryotic genome, comprising
CC (a) algorithmically predicting at least one exon from genomic sequences
CC of the eukaryote; and (b) detecting specific hybridisation of detectably
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC having a fragment identical to the predicted exon, the probe is included
CC in the above mentioned microarray; assigning exons to a single gene,
CC comprising (a) identifying exons from genomic sequence by the method
CC above and (b) measuring the expression of each of the exons in several
CC tissues and/or cell types using hybridisation to a single exon
CC microarrays having a probe with the exon, where a common pattern of
CC expression of the exons in the tissues and/or cell types indicates that
CC the exons should be assigned to a single gene; a peptide comprising one
CC of 12011 sequences, mentioned in the specification, or encoded by the
CC probes/open reading frames (ORF). The probes are used for gene
CC expression analysis, and for identifying exons in a gene, particularly
CC using human lung derived mRNA and for the study of lung diseases
CC such as asthma, lung cancer, chronic obstructive pulmonary disease
CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary

```

CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
 CC haemoidrosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
 CC pulmonary alveolar proteinosis, Karagenen syndrome, fibrocystic
 CC and hyaline membrane disease. The present sequence is a peptide/protein
 CC encoded by a single exon probe of the invention.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SQ Sequence 66 AA;

Query Match 2.8%; Score 12; DB 23; Length 66;

Best Local Similarity 100.0%; Pred. No. 0.0017;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

310 AHLRWHTGERPF 321

|||||
 30 AHLRWHTGERPF 41

RESULT 8

ABP48203

ID ABP48203 standard; Peptide; 90 AA.

AC ABP48203;

XX

DT 28-AUG-2002 (first entry)

XX

DE Sp-1 zinc finger protein SEQ ID NO:15.

XX

KW Zinc finger protein; ZFP; DNA binding protein; zinc finger.

XX

OS Synthetic.

XX

PN WO200242459-A2.

XX

PD 30-MAY-2002.

XX

PF 20-NOV-2001; 2001WO-US43438.

XX

PR 20-NOV-2000; 2000US-0716637.

XX

(SANG-) SANGAMO BIOSCIENCES INC.

XX

PI Liu Q;

XX

DR WPI; 2002-500284/53.

XX

PT New zinc finger protein that binds to target site, useful in studying

PT gene function and for human therapeutics and plant engineering,

PT comprises first, second and third zinc fingers, ordered from N- to

PT C-terminus

XX

PS Disclosure; Page 17; 81pp; English.

XX

CC The present invention describes a zinc finger protein (I) that binds to
 CC a target site, comprising a first (F1), a second (F2), and a third (F3)
 CC zinc finger, ordered F1, F2, F3 from N-terminus to C-terminus, where the
 CC target site comprises, in 3'-5' direction, a first (S1), a second (S2),
 CC and a third (S3) target site. Also described are: (1) a polypeptide
 CC (II) comprising (I); (2) a polynucleotide (III) encoding (I) or (II); and
 CC (3) designing (M) (I) involves selecting the F1 zinc finger such that
 CC it binds to the S1 target site, selecting the F2 zinc finger such
 CC that it binds to the S2 target site, and selecting the F3 zinc
 CC finger such that it binds to the S3 target site, thus designing (I)
 CC that binds to a target site. (I) is useful for recognition of triplet
 CC target sites having the nucleotide G in the 5'-most position of the
 CC target site. (I) is useful in studying gene function, and for human
 CC therapeutics and plant engineering. (I), (II) or (III) is useful in
 CC therapeutic methods to modulate the expression of a target region within

CC a subject, in diagnostic methods for sequence specific detection of
 CC target nucleic acid in a sample, and in assays to determine the
 CC phenotype and function of gene expression. (I) has improved affinity
 CC and specificity for their target sequences, as well as enhanced
 CC biological activity. AB071213 to AB072214 and AB048191 to AB051230
 CC represent DNA target sequences and zinc finger peptides which are given
 CC in the exemplification of the present invention.
 XX
 SQ Sequence 90 AA;

Query Match 2.8%; Score 12; DB 23; Length 90;

Best Local Similarity 100.0%; Pred. No. 0.0023;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 RSDHLSKHQRTH 373

|||||

Db 74 RSDHLSKHQRTH 85

RESULT 9

AAB14293

ID AAB14293 standard; peptide; 94 AA.

XX

AC AAB14293;

XX

DT 08-MAR-2001 (first entry)

XX

DE Transcription factor Sp-1 protein fragment #1.

XX

KW Transcription factor; Sp-1; DNA binding domain; bacterial;

KW viral; infection; oncogene; tumour suppressor gene; growth factor;

KW cancer; CCR5 receptor; HIV-1; apoE; Alzheimer's disease;

KW foetal haemoglobin; sickle cell anaemia; zinc finger protein.

XX

OS Unidentified.

XX

PN WO200042219-A1.

XX

PD 20-JUL-2000.

XX

PF 06-JAN-2000; 2000WO-US00388.

XX

PR 12-JAN-1999; 99US-0229007.

XX

(SANG-) SANGAMO BIOSCIENCES INC.

XX

PI Eisenberg SP, Case CC, Cox GN, Jamieson A, Rebar EJ;

XX

DR WPI; 2000-482840/42.

XX

PT Novel methods for selecting target sites for, and production of, zinc

PT finger proteins, useful for controlling expression of target genes,

PT e.g. for inhibiting oncogenes or treating sickle cell anemia

XX

PS Disclosure; Page 16; 83pp; English.

XX

CC Zinc finger proteins (ZFPs) bind to DNA and can modulate the expression
 CC of a wide range of genes. The present invention relates to a method for
 CC selecting a target site within a nucleic acid to be targeted by a ZFP, by
 CC detecting a specific 10-base motif. The present sequence is the DNA
 CC binding domain of transcription factor Sp-1. This sequence is a ZFP. This
 CC sequence was used as a framework for designing or selecting a
 CC nonnaturally occurring or variant ZFP. ZFPs can be designed to bind to
 CC preselected target genes to modulate their expression i.e. activate or
 CC inhibit. The target genes may be involved in disease e.g. bacterial or
 CC viral genes (involved in bacterial or viral infections); oncogenes,
 CC tumour suppressor genes or growth factors (involved in cancer); CCR5
 CC receptor gene (implicated in HIV-1 infection) or apoE gene (implicated
 CC in Alzheimer's disease). In addition, foetal haemoglobin genes may be
 CC induced by a variant ZFP for treating sickle cell anaemia. Furthermore,
 CC plant genes involved in herbicide or disease resistance may be used as
 CC target genes, to improve plant characteristics.

XX

SQ Sequence 94 AA;
 Query Match 2.8%; Score 12; DB 21; Length 94;
 Best Local Similarity 100.0%; Pred. No. 0.0024;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 310 AHLRWHTGERPF 321
 |||||
 Db 26 AHLRWHTGERPF 37

RESULT 10
 ABJ03953
 ID ABJ03953 standard; Peptide; 94 AA.
 AC ABJ03953;
 XX
 XX
 DT 25-SEP-2002 (first entry)
 XX
 DE Zinc finger protein Spl DNA binding domain SEQ ID NO: 221.
 XX
 KW Zinc finger protein; angiogenesis; vasculogenesis; ischaemia;
 XX diabetic retinopathy; psoriasis; arthropathy; cancer; tumour growth;
 XX gene therapy; antiatherosclerotic; vasotropic; antiarthritic; vulnerary;
 XX antitumor; cytostatic; antipsoriatic; antidiabetic; ophthalmological;
 KW osteopathic; antiinfertility.
 XX
 OS Unidentified.
 XX
 PN WO200246412-A2.
 XX
 PD 13-JUN-2002.
 XX
 PF 06-DEC-2001; 2001WO-US46861.
 XX
 PR 07-DEC-2000; 2000US-0733604.
 PR 12-DEC-2000; 2000US-0736083.
 PR 30-APR-2001; 2001US-0846033.
 XX
 PA (SANG-) SANGAMO BIOSCIENCES INC.
 XX
 PI Rebar E, Jamieson A, Liu Q, Liu P, Wolffe A, Eisenberg SP;
 PI Jarvis E;
 XX
 DR WPI; 2002-527918/56.
 XX
 PT New zinc finger protein that binds to target site in vascular
 PT endothelial growth factor gene, useful for modulating expression of the
 PT gene and for treating atherosclerosis, ischemia, arthritis, wound or
 PT ulcer
 XX
 XX Disclosure; Page 189; 195pp; English.

CC The present invention relates to a zinc finger protein that binds to a
 CC target site in one or more vascular endothelial growth factor (VEGF)
 CC genes. The protein is useful for modulating expression of a VEGF gene,
 CC thereby regulating angiogenesis and vasculogenesis. This can be used to
 CC treat atherosclerosis, ischaemia, arthritis, wounds, ulcers, tumours,
 CC diabetic retinopathy or psoriasis. The present sequence is a peptide
 CC shown in the invention.
 XX
 XX Sequence 94 AA;
 Query Match 2.8%; Score 12; DB 23; Length 94;
 Best Local Similarity 100.0%; Pred. No. 0.0024;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 310 AHLRWHTGERPF 321
 |||||
 Db 26 AHLRWHTGERPF 37

RESULT 11
 AAB14294
 ID AAB14294 standard; peptide; 100 AA.
 XX
 AC AAB14294;

ABP48202
 ID ABP48202 standard; Peptide; 94 AA.
 XX
 AC ABP48202;
 XX
 XX 28-AUG-2002 (first entry)
 DT
 DE Sp-1 zinc finger protein SEQ ID NO:13.
 XX
 KW Zinc finger protein; ZFP; DNA binding protein; zinc finger.
 XX
 OS Synthetic.
 XX
 PN WO200242459-A2.
 XX
 PD 30-MAY-2002.
 XX
 PF 20-NOV-2001; 2001WO-US43438.
 XX
 PR 20-NOV-2000; 2000US-0716637.
 XX
 PA (SANG-) SANGAMO BIOSCIENCES INC.
 XX
 PI Liu Q;
 XX
 DR WPI; 2002-500284/53.
 XX
 PT New zinc finger protein that binds to target site, useful in studying
 PT gene function and for human therapeutics and plant engineering,
 PT comprises first, second and third zinc fingers, ordered from N- to
 PT C-terminus
 XX
 PS Disclosure; Page 17; 81pp; English.

CC The present invention describes a zinc finger protein (I) that binds to
 CC a target site, comprising a first (F1), a second (F2), and a third (F3)
 CC zinc finger, ordered F1, F2, F3 from N-terminus to C-terminus, where the
 CC target site comprises, in 3'-5' direction, a first (S1), a second (S2),
 CC and a third (S3) target subsite. Also described are: (1) a polypeptide
 CC (II) comprising (I); (2) a polynucleotide (III) encoding (I) or (II); and
 CC (3) designing (M) (I) involves selecting the F1 zinc finger such that
 CC it binds to the S1 target subsite, selecting the F2 zinc finger such that
 CC that it binds to the S2 target subsite, and selecting the F3 zinc
 CC finger such that it binds to the S3 target subsite, thus designing (I)
 CC that binds to a target site. (I) is useful for recognition of triplet
 CC target subsites having the nucleotide G in the 5'-most position of the
 CC subsite. (I) is useful in studying gene function, and for human
 CC therapeutics and plant engineering. (I), (II) or (III) is useful in
 CC therapeutic methods to modulate the expression of a target region within
 CC a subject, in diagnostic methods for sequence specific detection of
 CC target nucleic acid in a sample, and in assays to determined the
 CC phenotype and function of gene expression. (I) has improved affinity
 CC and specificity for their target sequences, as well as enhanced
 CC biological activity. AB071213 to AB072214 and ABP48191 to ABP51230
 CC represent DNA target sequences and zinc finger peptides which are given
 CC in the exemplification of the present invention.
 XX
 SQ Sequence 94 AA;
 Query Match 2.8%; Score 12; DB 23; Length 94;
 Best Local Similarity 100.0%; Pred. No. 0.0024;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 310 AHLRWHTGERPF 321
 |||||
 Db 26 AHLRWHTGERPF 37

RESULT 12
 AAB14294
 ID AAB14294 standard; peptide; 100 AA.
 XX
 AC AAB14294;

```

XX DT 08-MAR-2001 (first entry)
XX DE Transcription factor Sp-1 protein fragment #2.
XX KW Transcription factor; Sp-1; DNA binding domain; bacterial;
XX KW viral; infection; oncogene; tumour suppressor gene; growth factor;
XX KW cancer; CCR5 receptor; HIV-1; apoE; Alzheimer's disease;
XX KW foetal haemoglobin; sickle cell anaemia; zinc finger protein.
XX OS Unidentified.
XX FH Key Location/Qualifiers
XX FT Peptide 1..10
XX FT /label= Leader_sequence
XX W0200042219-A1.
XX 20-JUL-2000.
XX PF 06-JAN-2000; 2000WO-US00388.
XX PR 12-JAN-1999; 99US-0229007.
XX PA (SANG-) SANGAMO BIOSCIENCES INC.
XX PI Eisenberg SP, Case CC, Cox GN, Jamieson A, Rebar EJ;
XX WPI; 2000-482840/42.
XX PT Novel methods for selecting target sites for, and production of, zinc
XX PT finger proteins, useful for controlling expression of target genes,
XX PT e.g. for inhibiting oncogenes or treating sickle cell anaemia
XX PS Disclosure; Page 16; 83pp; English.
XX CC Zinc finger proteins (ZFPs) bind to DNA and can modulate the expression
XX CC of a wide range of genes. The present invention relates to a method for
XX CC selecting a target site within a nucleic acid to be targeted by a ZFP, by
XX CC detecting a specific 10-base motif. The present sequence is the DNA
XX CC binding domain of transcription factor Sp-1. This sequence is a ZFP. This
XX CC sequence was used as a framework for designing or selecting a
XX CC nonnaturally occurring or variant ZFP. ZFPs can be designed to bind to
XX CC preselected target genes to modulate their expression i.e. activate or
XX CC inhibit. The target genes may be involved in disease e.g. bacterial or
XX CC viral genes (involved in bacterial or viral infections); oncogenes,
XX CC tumour suppressor genes or growth factors (involved in cancer); CCR5
XX CC receptor gene (implicated in HIV-1 infection) or apoE gene (implicated
XX CC in Alzheimer's disease). In addition, foetal haemoglobin genes may be
XX CC induced by a variant ZFP for treating sickle cell anaemia. Furthermore,
XX CC plant genes involved in herbicide or disease resistance may be used as
XX CC target genes, to improve plant characteristics.
XX SQ Sequence 100 AA;
XX Query Match 2.8%; Score 12; DB 21; Length 100;
XX Best Local Similarity 100.0%; Pred. No. 0.0025;
XX Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX QY 362 RSDHLSKHQRTH 373
XX DB 84 RSDHLSKHQRTH 95
XX RESULT 13
XX ABJ03954
XX ID ABJ03954 standard; Peptide; 100 AA.
XX AC ABJ03954;
XX XX 25-SEP-2002 (first entry)
XX DE Zinc finger protein Spl DNA binding domain SEQ ID NO: 223.

```

```

XX KW zinc finger protein; angiogenesis; vasculogenesis; ischaemia;
XX KW diabetic retinopathy; psoriasis; arthropathy; cancer; tumour growth;
XX KW gene therapy; antiatherosclerotic; vasotropic; antiarthritic; vulnery;
XX KW antiulcer; cytostatic; antipsoriatic; antidiabetic; ophthalmological;
XX KW osteopathic; antifertility.
XX OS Unidentified.
XX PN W0200246412-A2.
XX PD 13-JUN-2002.
XX PF 06-DEC-2001; 2001WO-US46861.
XX PR 07-DEC-2000; 2000US-0733604.
XX PR 12-DEC-2000; 2000US-0736083.
XX PR 30-APR-2001; 2001US-0846033.
XX PA (SANG-) SANGAMO BIOSCIENCES INC.
XX PI Rebar E, Jamieson A, Liu Q, Liu P, Wolffe A, Eisenberg SP;
XX PI Jarvis E;
XX XX WPI; 2002-527918/56.
XX PT New zinc finger protein that binds to target site in vascular
XX PT endothelial growth factor gene, useful for modulating expression of the
XX PT gene and for treating atherosclerosis, ischemia, arthritis, wound or
XX PT ulcer
XX PS Disclosure; Page 189; 195pp; English.
XX CC The present invention relates to a zinc finger protein that binds to a
XX CC target site in one or more vascular endothelial growth factor (VEGF)
XX CC genes. The protein is useful for modulating expression of a VEGF gene,
XX CC thereby regulating angiogenesis and vasculogenesis. This can be used to
XX CC treat atherosclerosis, ischaemia, arthritis, wounds, ulcers, tumours,
XX CC diabetic retinopathy or psoriasis. The present sequence is a peptide
XX CC shown in the invention.
XX SQ Sequence 100 AA;
XX Query Match 2.8%; Score 12; DB 23; Length 100;
XX Best Local Similarity 100.0%; Pred. No. 0.0025;
XX Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX QY 362 RSDHLSKHQRTH 373
XX DB 84 RSDHLSKHQRTH 95
XX RESULT 14
XX ABP41747
XX ID ABP41747 standard; Protein; 135 AA.
XX AC ABP41747;
XX XX 22-AUG-2002 (first entry)
XX DE Human ovarian antigen HSQEK12, SEQ ID NO:2879.
XX KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
XX KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
XX KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
XX KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
XX KW inflammatory condition; immune disorder; blood disorder;
XX KW cardiovascular disorder; respiratory disorder; neurological disorder;
XX KW gastrointestinal disorder; urinary system disorder; drug screening;
XX KW gene therapy; chromosome mapping; forensic analysis;
XX KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
XX KW antiinflammatory; gynaecological; reproductive.

```

```

OS XX Homo sapiens.
PN XX WO200200677-A1.
XX PD 03-JAN-2002.
XX PF 07-JUN-2001; 2001WO-US18569.
XX PR 07-JUN-2000; 2000US-209467P.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Birse CE, Rosen CA;
XX DR WPI; 2002-147878/19.
XX DR N-PSDB; ABQ34824.
XX PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,
XX PT useful in the prevention, treatment and diagnosis of cancer (e.g.
XX PT ovarian cancer), immune disorders, cardiovascular disorders and
XX PT neurological diseases -
XX PT
XX Claim 11; SEQ ID No 2879; 2922pp; English.
XX
XX The invention relates to 2175 novel human ovarian antigens (ABP41054-
XX CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56303), and also
XX CC encompasses polypeptides 90% identical and polynucleotides 95% identical
XX CC to the sequences of the invention. The invention additionally relates to
XX CC recombinant vectors and host cells comprising human ovarian antigen
XX CC polynucleotides, antibodies against human ovarian antigens, and the use
XX CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
XX CC treating, prognosing or preventing various ovary and/or breast-related
XX CC disorders. Such conditions include ovarian cancer and breast cancer, and
XX CC metastatic tumours of ovarian or breast origin, reproductive system
XX CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
XX CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
XX CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
XX CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
XX CC vaginitis), immune disorders (e.g., congenital and acquired
XX CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
XX CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
XX CC respiratory disorders, neurological disorders, gastrointestinal disorders
XX CC and urinary system disorders. Ovarian antigen polypeptides and
XX CC polynucleotides may also be used in screening for compounds which
XX CC modulate ovarian antigen expression or activity. The polynucleotides may
XX CC further be used for gene therapy, chromosome mapping, in the
XX CC identification of individuals and in forensic analysis, and the
XX CC polypeptides may be used as food additives or to prepare antibodies
XX CC useful in disease diagnosis, drug targeting and phenotyping. The present
XX CC sequence represents a human ovarian antigen of the invention.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 135 AA;
Query Match 2.8%; Score 12; DB 23; Length 135;
Best Local Similarity 100.0%; Pred. No. 0.0033;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 327 FCGRRTSRDEL 338
| | | | | | | | | |
DB 83 FCGRRTSRDEL 94
| | | | | | | | | |
RESULT 15
AAR95975
ID AAR95975 standard; peptide; 168 AA.
XX AAR95975;
XX
XX 19-FEB-1997 (first entry)
XX

```

DE XX Nucleic acid recognition unit #11.

KW Target binding assembly; nucleic acid recognition unit; NF-kappa-B; SPI;
TATA; human papillomavirus; HPV; HPV E2; human immunodeficiency virus;
KW HIV LTR; Tat binding unit; probe; assembly sequence; asymmetry sequence;
KW nuclear localisation signal sequence; human cell; HPV LTR; HIV; bacteria;
virus.

XX Synthetic.

OS WO9617956-A2.

PN 13-JUN-1996.

PD 07-DEC-1995; 95WO-US15944.

PR 09-DEC-1994; 94US-0353476.

XX (GENE-) GENE POOL INC.

PA Weininger AM, Weininger S;

XX WPI; 1996-287199/29.

XX Probe nucleic acids, target binding assemblies, etc - for detection
and localisation of specific nucleic acid sequences, esp. HIV and
HPV

PS Claim 14; Page 97-98; 172pp; English.

XX AAR95965-R55993 represent the nucleic acid recognition units (NAR) of
target binding assemblies (TBA) of the invention. These NARs are
selected from NF-kappa-B, SPI, TATA, human papillomavirus (HPV) E2, HPV
LTR, human immunodeficiency virus (HIV) LTR and Tat binding units. This
sequence represents a SPI sequence NAR. The TBA is recognised by the
target binding region (TBR) of a probe of the invention. The probe of
the invention contains a TBR, a booster binding region (BBR), and an
optional support or attachment (OSA). The TBA contains at least one
NAR, and optionally a linker sequence, an assembly sequence, an
asymmetry sequence, a nuclear localisation signal sequence, and an OSA.
The assembly sequence and asymmetry sequences are responsible for the
folding and association of the NARs. The linker sequence is an
oligopeptide, which does not interfere with NAR function, but provides
stability and control over the spacing of the NAR from the rest of the
TBA. The OSA is an attached support or indicator, or other means of
localisation of the probe. The probe can be used in a method for
detecting or localising a specific target nucleic acid sequence (TNA).
The method is highly sensitive, and has a high degree of specificity.
The method can be used for detecting specific nucleic acid sequences,
including those found in human cells, in HIV, HPV, and other nucleic acid
containing systems, including bacteria and viruses.

XX Sequence 168 AA;

Query Match 2.8%; Score 12; DB 17; Length 168;
Best Local Similarity 100.0%; Pred. No. 0.004;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 310 AHLRWHTGERPF 321

| | | | | | | | | |
DB 28 AHLRWHTGERPF 39

Search completed: February 19, 2003, 13:37:41
Job time : 41 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model
Run on: February 19, 2003, 13:36:42 ; Search time 21 Seconds
(without alignments)
1959.311 Million cell updates/sec

Title: US-09-734-329-2
Perfect score: 428
Sequence: 1 MASSLLEEAHYGSSPLAML.....PAPPEKXHGSPQSNLLEI 428

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

...rched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR73:*
1: Pirl1.*
2: Pirl2.*
3: Pirl3.*
4: Pirl4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	12	2.8	495	A4489	GT box-binding pro
2	12	2.8	696	A29635	transcription fact
3	12	2.8	697	B4489	GT box-binding pro
4	12	2.8	780	A48143	HF-1-regulatory el
5	12	2.8	784	S26638	SPR-1 protein - hu
6	12	2.8	788	J50747	regulatory protein
7	11	2.6	644	S39356	transcription fact
8	10	2.3	298	T22250	hypothetical prote
9	10	2.3	1263	T13805	spalt-related prot
10	9	2.1	139	S06550	finger protein (cl
11	9	2.1	200	T36715	hypothetical prote
12	9	2.1	207	T50206	hypothetical zinc
13	9	2.1	338	G01436	transcription fact
14	9	2.1	348	I38599	zinc finger protei
15	9	2.1	363	I38937	DNA/RNA-binding pr
16	9	2.1	485	A40751	finger protein MZF
17	8	1.9	101	S30493	Spl protein - mous
18	8	1.9	219	S35643	BREB2 protein - hu
19	8	1.9	224	B72527	hypothetical prote
20	8	1.9	244	I59602	GC box binding pro
21	8	1.9	244	J50748	basic transcriptio
22	8	1.9	244	S25288	BREB protein - rat
23	8	1.9	250	T46417	hypothetical prote
24	8	1.9	318	H89754	two-component sens
25	8	1.9	362	T45072	erythroid Kruppel-
26	8	1.9	376	A48060	erythroid Kruppel-
27	8	1.9	416	1 FOCH	transforming prote
28	8	1.9	416	1 TVEFAC	transforming prote
29	8	1.9	423	1 TVFV2E	transforming prote

ALIGNMENTS

RESULT 1

A44489

GT box-binding protein Sp2 - human

N:Alternate names: transcription factor Sp2

C:Species: Homo sapiens (man)

C>Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999

C:Accession: A44489

R:Kingsley, C.; Winoto, A.

Mol. Cell. Biol. 12, 4251-4261, 1992

A:Title: Cloning of GT box-binding proteins: a novel Spl multigene family regulating

A:Reference number: A44489; MUID:93024366; PMID:1341900

A:Accession: A44489

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-495 <KIN>

A:Cross-references: GB:M97190; NID:g3383300; PIDN:AAA36629.1; PID:g3383301

A:Experimental source: Molt13

A>Note: sequence extracted from NCBI backbone (NCBIP:114078)

C:Keywords: DNA binding; transcription regulation

Query Match 2.8% Score 12; DB 2; Length 495;

Best Local Similarity 100.0%; Pred. No. 0.0018;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 327 FCGKRFTRSDDEL 338

|||||

Db 443 FCGKRFTRSDDEL 454

RESULT 2

A29635

transcription factor Spl - human (fragment)

N:Alternate names: finger protein ZNF76

C:Species: Homo sapiens (man)

C>Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 05-Nov-1999

C:Accession: A29635; G44256

R:Kadonaga, J.T.; Carner, K.R.; Maslarsz, F.R.; Tjian, R.

Cell 51, 1079-1090, 1987

A:Title: Isolation of cDNA encoding transcription factor Spl and functional analysis

A:Reference number: A29635; MUID:88080466; PMID:3319186

A:Accession: A29635

A:Molecule type: mRNA

A:Residues: 1-696 <KAD>

A:Cross-references: GB:J03133; NID:g339517; PIDN:AAA61154.1; PID:g339518

R:Ragoussis, J.; Senger, G.; Mockridge, I.; Sanseau, P.; Ruddy, S.; Dudley, K.; Shee

Genomics 14, 673-679, 1992

A:Title: A testis-expressed Zn finger gene (ZNF76) in human 6p21.3 centromeric to t(

A:Reference number: A44256; MUID:93052398; PMID:1427894

A:Accession: G44256

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 569-598 <RAG>

A:Experimental source: T-cell line CEM
A:Note: sequence extracted from NCBI backbone (NCBIP:125980)

C:Genetics:

A:Gene: GDB:SP1

A:Cross-references: GDB:127453; OMIM:189906

A:Map position: 19q13.1-19q13.3

C:Keywords: DNA binding; transcription regulation; zinc finger

Query Match 2.8%; Score 12; DB 2; Length 696;

Best Local Similarity 100.0%; Pred. No. 0.0024;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 310 AHLRWHTGERPF 321

Db 556 AHLRWHTGERPF 567

RESULT 3

B4489

GT box-binding protein Sp3 - human (fragment)

N:Alternate names: SPR-2 protein; transcription factor Sp3

C:Species: Homo sapiens (man)

A:Date: 10-Jun-1993 #sequence_revision 10-May-1996 #text_change 08-Oct-1999

A:Accession: B4489; S26639

A:Gene: GDB:SP4; SPR-1

A:Cross-references: GDB:136781

A:Map position: 2q31-2q31

C:Keywords: DNA binding; transcription regulation

Query Match 2.8%; Score 12; DB 2; Length 784;

Best Local Similarity 100.0%; Pred. No. 0.0027;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 310 AHLRWHTGERPF 321

Db 666 AHLRWHTGERPF 677

RESULT 6

JS0747

regulatory protein Spl - rat

C:Species: Rattus norvegicus (Norway rat)

A:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 01-Dec-2000

A:Accession: JS0747; S25287

R:Inataka, H.; Sogawa, K.; Yasumoto, K.; Kikuchi, Y.; Sasano, K.; Kobayashi, A.; Hay

submitted to JIPID, September 1992

A:Reference number: JS0747

A:Accession: JS0747

A:Molecule type: mRNA

A:Residues: 1-788 <INA>

A:Cross-references: DDBJ:D12768; NID:g220911; PIDN:CAA02235.1; PID:d1002730; PTD:g22

R:Inataka, H.; Sogawa, K.; Yasumoto, K.; Kikuchi, Y.; Sasano, K.; Kobayashi, A.; Hay

EMBO J. 11, 3663-3671, 1992

A:Title: Two regulatory proteins that bind to the basic transcription element (BTE),

A:Reference number: S25287; MUID:93010958; PMID:1356762

A:Accession: S25287

A:Molecule type: mRNA

A:Residues: 1-122, 'L', 124-311, 'A', 313-788 <IM2>

C:Keywords: DNA binding; transcription regulation

Query Match 2.8%; Score 12; DB 2; Length 788;

Best Local Similarity 100.0%; Pred. No. 0.0027;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 310 AHLRWHTGERPF 321

Db 648 AHLRWHTGERPF 659

RESULT 7

S39356

transcription factor btd - fruit fly (Drosophila sp.)

C:Species: Drosophila sp.

A:Date: 18-Feb-1994 #sequence_revision 01-Dec-1995 #text_change 07-May-1999

C:Accession: S39356
R:Wimmer, E.A.; Jaeckle, H.; Pfeifle, C.; Cohen, S.M.
Nature 366, 690-694, 1993
A:Title: A Drosophila homologue of human Spl is a head-specific segmentation gene.
A:Reference number: S39356; MUID:94081952; PMID:8259212
A:Accession: S39356
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-844 <WIM>
A:Cross-references: EMBL:229361; NID:9441283; PID:9441284
C:Genetics:
A:Gene: FlyBase:btd
A:Cross-references: FlyBase:FBgn0000233
A:Introns: 245/2

Query Match 2.6%; Score 11; DB 2; Length 644;
Best Local Similarity 100.0%; Pred. No. 0.023;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 311 HLRWHTGERPF 321
|||||
Db 353 HLRWHTGERPF 363

RESULT 8
T22250
hypothetical protein F45H11.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T22250
R:Kelly, P.
submitted to the EMBL Data Library, August 1996
A:Reference number: Z19537
A:Accession: T22250
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-298 <WIL>
A:Cross-references: EMBL:278420; PIDN:CAB01709.1; GSPDB:GN00019; CESP:F45H11.1
A:Experimental source: clone F45H11
C:Genetics:
A:Gene: CESP:F45H11.1
A:Map position: 1
A:Introns: 52/3; 109/2; 177/3; 248/1; 290/3

Query Match 2.3%; Score 10; DB 2; Length 298;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 306 SHLKAHLRW 315
|||||
Db 275 SHLKAHLRW 284

RESULT 9
T13805
spalt-related protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13805
R:Barrio, R.; Shea, M.J.; Carulli, J.; Lipkow, K.; Gaul, U.; Frommer, G.; Schuh, R.; Jac
submitted to the EMBL Data Library, August 1996
A:Description: The spalt-related gene of Drosophila melanogaster is a member of an ancie
A:Reference number: Z17768
A:Accession: T13805
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1263 <BAR>
A:Cross-references: EMBL:Y07653; NID:e1004366; PID:e267584; PIDN:CAA68937.1
C:Genetics:
A:Cross-references: EMBL:Y07653; NID:e1004366; PID:e267584; PIDN:CAA68937.1
A:Introns: 862/1; 915/3; 1003/1; 1080/3; 1146/3

Query Match 2.3%; Score 10; DB 2; Length 1263;

Best Local Similarity 100.0%; Pred. No. 0.4;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 315 HTGERPFVCN 324
|||||
Db 343 HTGERPFVCN 352

RESULT 10
S06550
finger protein (clone XlCOF14) - African clawed frog (fragment)
C:Species: Xenopus laevis (African clawed frog)
C:Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 31-Dec-1993
C:Accession: S06550
R:Nietfeld, W.; El-Baradi, T.; Mentzel, H.; Pieler, T.; Koester, M.; Posting, A.; Ki
J. Mol. Biol. 208, 639-659, 1989
A:Title: Second-order repeats in Xenopus laevis finger proteins.
A:Reference number: S05632; MUID:90040698; PMID:2509712
A:Accession: S06550
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-139 <NIE>
C:Keywords: DNA binding; zinc finger

Query Match 2.1%; Score 9; DB 2; Length 139;
Best Local Similarity 100.0%; Pred. No. 0.58;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 315 HTGERPFVC 323
|||||
Db 111 HTGERPFVC 119

RESULT 11
T26715
hypothetical protein Y38H8A.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T26715
R:McMurray, A.
submitted to the EMBL Data Library, January 1998
A:Reference number: Z20256
A:Accession: T26715
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-200 <WIL>
A:Cross-references: EMBL:AL021483; PIDN:CAA16344.1; GSPDB:GN00022; CESP:Y38H8A.5
A:Experimental source: clone Y38H8A
C:Genetics:
A:Gene: CESP:Y38H8A.5
A:Map position: 4
A:Introns: 19/1; 59/3

Query Match 2.1%; Score 9; DB 2; Length 200;
Best Local Similarity 100.0%; Pred. No. 0.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 315 HTGERPFVC 323
|||||
Db 50 HTGERPFVC 58

RESULT 12
T50206
hypothetical zinc finger protein [imported] - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jun-2000
C:Accession: T50206
R:Barrell, B.G.; Rajadream, M.A.; McDougall, R.C.; McLean, J.; Harris, D.
submitted to the EMBL Data Library, November 1999
A:Reference number: Z25045
A:Accession: T50206
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-207 <BAR>
A:Cross-references: EMBL:AL133225; PIDN:CAB61785.1; GSPDB:GN00066; SPDB:SPAC25B8.19c
A:Experimental source: strain 972h(-); cosmid c25B8
C:Genetics:
A:Gene: SPDB:SPAC25B8.19c
A:Map position: 1

Query Match 2.1%; Score 9; DB 2; Length 207;
Best Local Similarity 100.0%; Pred. No. 0.82;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 315 HTGERPFVC 323
|||||
DB 175 HTGERPFVC 183

RESULT 13
G01496
transcription factor IIIA - human (fragment)
N:Alternate names: TFIIA
C:Species: Homo sapiens (man)
Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 05-Nov-1999
Accession: G01496
Becker, K.G.
submitted to the EMBL Data Library, August 1994
A:Reference number: G07520
A:Accession: G01496
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-338 <REC>
A:Cross-references: EMBL:U14134; NID:g551534; PIDN:AAA21873.1; PID:g551535
C:Genetics:
A:Gene: GDB:GTF3A; TFIIA
A:Cross-references: GDB:434744; OMIM:600860
A:Map position: 13q12.3-13q13.1
A:Superfamily: transcription factor IIIA

Query Match 2.1%; Score 9; DB 2; Length 338;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 315 HTGERPFVC 323
|||||
DB 62 HTGERPFVC 70

RESULT 14
I38599
zinc finger protein ZNF134 - human
Species: Homo sapiens (man)
Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 05-Nov-1999
Accession: I38599
R:Tommerup, N.; Vissing, H.
Genomics 27, 259-264, 1995
A:Title: Isolation and fine mapping of 16 novel human zinc finger-encoding cDNAs identified
A:Reference number: AS7785; MUID:96044430; PMID:7557990
A:Accession: I38599
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-348 <RES>
A:Cross-references: EMBL:U09412; NID:g488552; PIDN:AAC50253.1; PID:g488553
C:Genetics:
A:Gene: GDB:ZNF134
A:Cross-references: GDB:137033
A:Map position: 19q13.4-19q13.4

Query Match 2.1%; Score 9; DB 2; Length 348;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 315 HTGERPFVC 323
|||||

Db 287 HTGERPFVC 295

RESULT 15
I38937
DNA/RNA-binding protein - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jul-2000
C:Accession: I38937
R:Drew, P.D.; Nagle, J.W.; Canning, R.D.; Ozato, K.; Biddison, W.E.; Becker, K.G.
Gene 159, 215-218, 1995
A:Title: Cloning and expression analysis of a human cDNA homologous to Xenopus TFIIF
A:Reference number: I38937; MUID:95347600; PMID:7622052
A:Accession: I38937
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-363 <RES>
A:Cross-references: EMBL:U20272; NID:g644870; PIDN:AAA75623.1; PID:g644871
C:Superfamily: transcription factor IIIA

Query Match 2.1%; Score 9; DB 2; Length 363;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 315 HTGERPFVC 323
|||||
DB 62 HTGERPFVC 70

Search completed: February 19, 2003, 13:40:11
Job time : 22 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 19, 2003, 13:33:06 ; Search time 14 Seconds
(without alignments)
1267.991 Million cell updates/sec

Title: US-09-734-329-2

Perfect score: 428

Sequence: 1 MASSLEEHAHYGSSPLAML.....PAPPEKAHGGSPQSNLLEI 428

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

--arched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40:**

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	4.7	152	1	SP6_MOUSE
2	12	2.8	252	1	BTEA_HUMAN
3	12	2.8	606	1	SP2_HUMAN
4	12	2.8	711	1	SP3_HUMAN
5	12	2.8	784	1	SP4_HUMAN
6	12	2.8	785	1	SP1_HUMAN
7	12	2.8	788	1	SP1_RAT
8	11	2.6	644	1	BTD_DROME
9	9	2.1	139	1	Z014_XENLA
10	9	2.1	251	1	BTE4_MOUSE
11	9	2.1	348	1	Z134_HUMAN
12	9	2.1	423	1	TF3A_HUMAN
13	9	2.1	734	1	ZN42_HUMAN
14	8	1.9	244	1	BTE1_HUMAN
15	8	1.9	244	1	BTE1_MOUSE
16	8	1.9	244	1	BTE1_RAT
17	8	1.9	288	1	KLFD_HUMAN
18	8	1.9	289	1	KLFD_MOUSE
19	8	1.9	351	1	KLF2_RAT
20	8	1.9	354	1	KLF2_MOUSE
21	8	1.9	355	1	KLF2_HUMAN
22	8	1.9	358	1	KLF1_MOUSE
23	8	1.9	362	1	KLF1_HUMAN
24	8	1.9	415	1	KLFF_MOUSE
25	8	1.9	416	1	KLFF_HUMAN
26	8	1.9	416	1	MYC_AVIOR
27	8	1.9	416	1	MYC_CHICK
28	8	1.9	421	1	MYC_AVID
29	8	1.9	422	1	MYC_AVIDC
30	8	1.9	422	1	MYC_AVIDM
31	8	1.9	423	1	MYC_AVIDE
32	8	1.9	445	1	KLF5_MOUSE
33	8	1.9	457	1	KLF5_HUMAN

34 8 1.9 470 1 KLF4_HUMAN O43474 homo sapien
35 8 1.9 474 1 KLF4_MOUSE Q60793 mus musculus
36 8 1.9 494 1 CRTD_RHOCA P17059 rhodobacter
37 8 1.9 512 1 KLF8_HUMAN O14901 homo sapien
38 8 1.9 1043 1 PTF1_DROME P33244 drosophila
39 7 1.6 128 1 CSH3_MOUSE P97803 mus musculus
40 7 1.6 139 1 ZG67_XENLA P18734 xenopus lae
41 7 1.6 140 1 ZG49_XENLA P18724 xenopus lae
42 7 1.6 168 1 ZG7_XENLA P18735 xenopus lae
43 7 1.6 168 1 ZG7_XENLA P18742 xenopus lae
44 7 1.6 194 1 TRPG_HELPJ Q92ju6 helicobacte
45 7 1.6 194 1 TRPG_HELPJ Q25868 helicobacte

ALIGNMENTS

RESULT 1
SP6_MOUSE STANDARD; PRT; 152 AA.
AC Q9ESX2;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Transcription factor Sp6 (Krueppel-like factor 14) (Fragment).
GN SP6 OR KLF14.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20541714; PubMed=11087666;
RA Scohy S., Gabant P., Van Reeth T., Hertveldt V., Dreze P.-L.,
RA Van Vooren P., Riviere M., Spierer J., Szpirer C.;
RT Identification of KLF13 and KLF14 (SP6), novel members of the SP/XKLF
transcription factor family.;
RL Genomics 70:93-101(2000).
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC -!- SIMILARITY: BELONGS TO THE SP1 FAMILY OF C2H2-TYPE ZINC-FINGER
PROTEINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AJ275988; CAC06698.1;
HSP; P08047; 1SP2.
MGD; MGI:1932575; Sp6.
InterPro; IPR000822; Znf_C2H2.
Pfam; PF000096; zf-C2H2; 3.
ProDom; PD000003; Znf_C2H2; 1.
SMART; SM00355; Znf_C2H2; 3.
PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
PROSITE; PS00157; ZINC_FINGER_C2H2_2; 3.
Transcription regulation; DNA-binding; Nuclear protein; Repeat;
Zinc-finger; Metal-binding.
NON_TER 1
DOMAIN 30 112 ZINC FINGERS.
ZN_FING 30 54 C2H2-TYPE.
ZN_FING 60 84 C2H2-TYPE.
ZN_FING 90 112 C2H2-TYPE.
SQ SEQUENCE 152 AA; 16456 MW; AF629C4845599938 CRC64;

Query Match 4.7%; Score 20; DB 1; Length 152;
Best Local Similarity 100.0%; Pred. No. 7.2e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 319 RPFVNCWLFCKGRTSRDEL 338
Db 58 RPFVNCWLFCKGRTSRDEL 77

RESULT 2
BTE4_HUMAN
ID BTE4_HUMAN STANDARD; PRT; 252 AA.
AC Q9BXK1;
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcription factor BTE4 (Basic transcription element binding-
DE protein 4) (BTE-binding protein 4) (Kruppel-like factor 16) (Novel
DE Spl-like zinc finger transcription factor 2) (Transcription factor
DE NSLP2).
GN KLF16 OR BTE4 OR NSLP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
TISSUE=Pancreas;
RA Conley A., Urrutia R.;
RT "Isolation of a novel zinc finger transcription factor from the
RT pancreas extends the repertoire of Spl-like proteins present in this
RT organ (Abstract #153).";
RL Pancreas 21:437-437(2000).
CC -!- FUNCTION: Transcription factor that binds GC and GT boxes and
CC displaces Spl and Sp3 from these sequences. Modulates dopaminergic
CC transmission in the brain (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SP1 FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AF327440; AAK15698.1; -
DR HSSP; P08047; ISP2.
DR GENE; HGNC:16857; KLF16.
DR MIM; 606139; -
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 3.
DR SMART; SM00335; Znf_C2H2; 3.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.
KW Transcription regulation; DNA-binding; Nuclear protein; Repeat;
KW Zinc-finger; Metal-binding.
FT DOMAIN 3 130 ALA/PRO-RICH.
FT DOMAIN 99 112 SER-RICH.
FT DOMAIN 127 209 ZINC FINGERS.
FT ZN_FING 127 150 C2H2-TYPE.
FT ZN_FING 157 181 C2H2-TYPE.
FT ZN_FING 187 209 C2H2-TYPE.
FT ZN_FING 210 249 PRO/SER-RICH.
FT DOMAIN 252 338
SQ SEQUENCE 252 AA; 25430 MW; 9A0CB4B1A585A118 CRC64;

Query Match 2.8%; Score 12; DB 1; Length 252;
Best Local Similarity 100.0%; Pred. No. 0.0064;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 355 LCSKRFTRSDHL 366
Db 191 LCSKRFTRSDHL 202

```

```

RESULT 3
SP2_HUMAN
ID SP2_HUMAN STANDARD; PRT; 606 AA.
AC Q02086;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcription factor Sp2.
DE Transcription factor Sp2.
GN SP2 OR KIAA0048.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
TISSUE=Bone marrow;
RC MEDLINE=96051398; PubMed=7584044;
RX Nomura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S.,
RA Seki N., Kawabayashi Y., Ishikawa K.-I., Tabata S.;
RT "Prediction of the coding sequences of unidentified human genes. II.
RT The coding sequences of 40 new genes (K1AA0041-K1AA0080) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 1:223-229(1994).
RN [2]
SEQUENCE OF 112-606 FROM N.A.
RX MEDLINE=93024366; PubMed=1341900;
RA Kingsley C., Winoto A.;
RT "Cloning of GT box-binding proteins: a novel Sp1 multigene family
RT regulating T-cell receptor gene expression.";
RL Mol. Cell. Biol. 12:4251-4261(1992).
CC -!- FUNCTION: BINDS TO GC BOX PROMOTERS ELEMENTS AND SELECTIVELY
CC ACTIVATES MNA SYNTHESIS FROM GENES THAT CONTAIN FUNCTIONAL
CC RECOGNITION SITES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE SP1 FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; D28588; BAA05923.1; -
DR EMBL; M97190; AAA36629.1; -
DR PIR; A44489; A44489.
DR HSSP; P08047; ISP2.
DR TRANSFAC; T02356; -
DR GENE; HGNC:11207; SP2.
DR MIM; 601801; -
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 3.
DR PRINTS; PR00048; ZINC_FINGER.
DR PRODOM; PD000003; Znf_C2H2; 2.
DR SMART; SM00355; Znf_C2H2; 3.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.
KW Transcription regulation; Activator; Zinc-finger; Metal-binding;
KW DNA-binding; Nuclear protein; Repeat.
FT DOMAIN 518 600 ZINC FINGERS.
FT ZN_FING 518 542 C2H2-TYPE.
FT ZN_FING 548 572 C2H2-TYPE.
FT ZN_FING 578 600 C2H2-TYPE.
FT SEQUENCE 606 AA; 64153 MW; A27C6D460D36E186 CRC64;

Query Match 2.8%; Score 12; DB 1; Length 606;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 327 FCGKRFTRSDHL 338
Db 191 FCGKRFTRSDHL 338

```

DB 554 PCGRFTRSDDEL 565

RESULT 4

SP3_HUMAN STANDARD; PRT; 711 AA.

AC 002447; 1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Transcription factor Sp3 (SPR-2) (Fragment).

GN SP3

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

[1]

SEQUENCE FROM N.A., AND REVISIONS.

TISSUE-T-cell;

RA Kingsley C., Winoto A.;

RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE OF 59-711 FROM N.A.

RX MEDLINE-93024366; PubMed=1341900;

RA Kingsley C., Winoto A.;

RT Cloning of CT box-binding proteins: a novel Sp1 multigene family

RT regulating T-cell receptor gene expression.;

RL Mol. Cell. Biol. 12:4251-4261(1992).

RN [3]

RP SEQUENCE OF 15-711 FROM N.A.

PC TISSUE-Uterus;

RX MEDLINE-93087156; PubMed=1454515;

RA Hagen G., Mueller S., Beato M., Suske G.;

RT Cloning by recognition site screening of two novel GT box binding

RT proteins: a family of Sp1 related genes.;

RL Nucleic Acids Res. 20:5519-5525(1992).

RN [4]

RP FUNCTION.

RX MEDLINE-97426517; PubMed=9278495;

RA Ihn H., Trojanowska M.;

RT "Sp3 is a transcriptional activator of the human alpha2(I) collagen

RT gene.;"

RL Nucleic Acids Res. 25:3712-3717(1997).

CC -!- FUNCTION: BINDS TO GT AND GC BOXES PROMOTERS ELEMENTS. PROBABLE

CC TRANSCRIPTIONAL ACTIVATOR.

CC -!- SUBCELLULAR LOCATION: Nuclear.

CC -!- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED.

CC -!- SIMILARITY: BELONGS TO THE SP1 FAMILY OF C2H2-TYPE ZINC-FINGER

CC PROTEINS.

CC

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

CC

EMBL; M97191; AAA36630.2; -

DR EMBL; X68560; CAA48562.1; -

DR PIR; S26639; S26639.

DR HSP; P08047; ISPI.

DR TRANSFAC; T02338; -

DR Genew; HGNC:11208; SP3.

DR MIN; 601804; -

DR InterPro; IPR000822; Znf_C2H2.

DR Pfam; PF00096; Zf-C2H2; 3.

DR ProDom; PD000003; Znf_C2H2; 1.

DR SMART; SM00355; Znf_C2H2; 3.

DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.

DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.

KW Transcription regulation; Activator; Zinc-finger; Metal-binding;

DNA-binding; Nuclear protein; Repeat.

FT NON_TER 1 1

FT DOMAIN 551 633 ZINC FINGERS.

FT ZN_FING 551 575 C2H2-TYPE.

FT ZN_FING 581 605 C2H2-TYPE.

FT ZN_FING 611 633 C2H2-TYPE.

FT CONFLICT 94 94 T -> A (IN REF. 3).

FT CONFLICT 669 669 N -> K (IN REF. 3).

SQ SEQUENCE 711 AA; 75102 MW; D1719FAC9B05A217 CRC64;

Query Match -2.8%; Score 12; DB 1; Length 711;

Best Local Similarity 100.0%; Pred. No. 0.0016;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 CHIPGCGKVGK 304

DB 553 CHIPGCGKVGK 564

|||||

RESULT 5

SP4_HUMAN STANDARD; PRT; 784 AA.

AC 002446; O60402;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Transcription factor Sp4 (SPR-1).

GN SP4.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-Uterus;

RX MEDLINE-93087156; PubMed=1454515;

RA Hagen G., Mueller S., Beato M., Suske G.;

RT Cloning by recognition site screening of two novel GT box binding

RT proteins: a family of Sp1 related genes.;

RL Nucleic Acids Res. 20:5519-5525(1992).

RN [2]

RP SEQUENCE FROM N.A.

RA Ozerksy P., Holmes A.;

RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: BINDS TO GT AND GC BOXES PROMOTERS ELEMENTS. PROBABLE

CC TRANSCRIPTIONAL ACTIVATOR.

CC -!- SUBCELLULAR LOCATION: Nuclear.

CC -!- TISSUE SPECIFICITY: ABUNDANT IN BRAIN.

CC -!- SIMILARITY: BELONGS TO THE SP1 FAMILY OF C2H2-TYPE ZINC-FINGER

CC PROTEINS.

CC

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

CC

EMBL; X68561; CAA48563.1; -

DR EMBL; AC004595; AAD12226.1; -

DR PIR; S26638; S26638.

DR HSP; P08047; ISPI.

DR TRANSFAC; T02339; -

DR Genew; HGNC:11209; SP4.

DR MIN; 600540; -

DR InterPro; IPR000822; Znf_C2H2.

DR Pfam; PF00096; Zf-C2H2; 3.

DR PRINTS; PR00048; ZINC_FINGER.

DR ProDom; PD000003; Znf_C2H2; 2.

DR SMART; SM00355; Znf_C2H2; 3.

DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.

DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.

KW Transcription regulation; Activator; Zinc-finger; Metal-binding;

KW DNA-binding; Nuclear protein; Repeat.
 FT DOMAIN 7 11 POLY-GLU.
 FT DOMAIN 12 19 POLY-ALA.
 FT DOMAIN 122 130 POLY-SER.
 FT DOMAIN 185 188 POLY-SER.
 FT DOMAIN 647 729 ZINC FINGERS.
 FT 2N_FING 647 729 C2H2-TYPE.
 FT 2N_FING 677 701 C2H2-TYPE.
 FT 2N_FING 707 729 C2H2-TYPE.
 FT CONFLICT 197 197 K -> Q (IN REF. 2).
 FT CONFLICT 379 380 HA -> QP (IN REF. 2).
 FT CONFLICT 386 386 Q -> A (IN REF. 2).
 SQ SEQUENCE 784 AA; 82025 MW; 3C4EAE28CB2B81FB CRC64;

Query Match 2.8%; Score 12; DB 1; Length 784;
 Best Local Similarity 100.0%; Pred. No. 0.0018;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 310 AHLRWHTGERPF 321
 |||||
 Db 666 AHLRWHTGERPF 677

ULT 6
 ID SPI_HUMAN STANDARD; PRT; 785 AA.
 AC P08047; Q9NVE7; Q9H3Q5;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Transcription factor Sp1.
 GN SPI OR TSFPI.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE OF 4-785 FROM N.A.
 RC TISSUE=Cervical carcinoma;
 RA Haggart M.H., Ladurner A.G.;
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 1-558 FROM N.A.
 RX MEDLINE=20545563; PubMed=10973950;
 RA Takahara T., Kanazu S., Yanagisawa S., Akanuma H.;
 RT "Heterogeneous Sp1 mRNAs in human HepG2 cells include a product of
 RT homotypic trans-splicing.";
 RL J. Biol. Chem. 275:38067-38072(2000).
 RN [3]
 RP SEQUENCE OF 90-785 FROM N.A., AND SEQUENCE OF 359-375 AND 670-675.
 PY MEDLINE=88080466; PubMed=3319186;
 RA Kadonaga J.T., Carner K.R., Maslarsz F.R., Tjian R.;
 RT "Isolation of cDNA encoding transcription factor Sp1 and functional
 RT analysis of the DNA binding domain.";
 RL Cell 51:1079-1090(1987).
 RN [4]
 RP O-GLYCOSYLATION.
 RA MEDLINE=89003041; PubMed=3139301;
 RX Jackson S.P., Tjian R.;
 RT "O-glycosylation of eukaryotic transcription factors: implications
 RT for mechanisms of transcriptional regulation.";
 RL Cell 55:125-133(1988).
 RN [5]
 RP STRUCTURE BY NMR OF 654-684 AND 684-712.
 RX MEDLINE=97218212; PubMed=9065444;
 RA Narayan V.A., Kriwacki R.W., Caradonna J.P.;
 RT "Structures of zinc finger domains from transcription factor Sp1.
 RT Insights into sequence-specific protein-DNA recognition.";
 RL J. Biol. Chem. 272:7801-7809(1997).
 RN [6]
 RP IDENTIFICATION OF SEROTONIN 1A-RECEPTOR PROMOTER BINDING SITES.
 RX MEDLINE=96224025; PubMed=8626793;
 RA Parks C.L., Shenk T.;

RT "The serotonin 1a receptor gene contains a TATA-less promoter that
 RT responds to MAZ and Sp1.";
 RL J. Biol. Chem. 271:4417-4430(1996).
 CC -|- FUNCTION: BINDS TO GC BOX PROMOTERS ELEMENTS AND SELECTIVELY
 CC ACTIVATES MRNA SYNTHESIS FROM GENES THAT CONTAIN FUNCTIONAL
 CC RECOGNITION SITES. CAN INTERACT WITH G/C-RICH MOTIFS FROM
 CC SEROTONIN RECEPTOR PROMOTER.
 CC -|- SUBCELLULAR LOCATION: Nuclear.
 CC -|- PTM: O-GLYCOSYLATED; CONTAINS N-ACETYLGLUCOSAMINE SIDE CHAINS.
 CC -|- SIMILARITY: BELONGS TO THE SP1 FAMILY OF C2H2-TYPE ZINC-FINGER
 CC PROTEINS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC -----

DR EMBL; AF252284; AAF67726.1; -;
 DR EMBL; AB039286; BAB13476.1; -;
 DR EMBL; J03133; AAA61154.1; -;
 DR PIR; A29635; A29635.
 DR PDB; 1SPI; 21-APR-97.
 DR PDB; 1SP2; 21-APR-97.
 DR TRANSFAC; T00759; -;
 DR GlycoSuiteDB; P08047; -;
 DR Genew; HGNC:11205; SP1.
 DR MIM; 189906; -;
 DR InterPro; IPR000822; Znf_C2H2.
 DR Pfam; PF000096; zf-C2H2; 3.
 DR ProDom; PD000003; Znf_C2H2; 2.
 DR SMART; SM00028; ZINC_FINGER_C2H2_1; 3.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
 DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 3.
 KW Transcription regulation; Activator; Zinc-finger; Metal-binding;
 KW DNA-binding; Nuclear protein; Repeat; Glycoprotein; 3D-structure.
 FT DOMAIN 626 708 ZINC FINGERS.
 FT 2N_FING 626 650 C2H2-TYPE.
 FT 2N_FING 656 680 C2H2-TYPE.
 FT 2N_FING 686 708 C2H2-TYPE.
 FT CONFLICT 366 366 D -> G (IN REF. 3; AA SEQUENCE).
 FT CONFLICT 670 670 S -> F (IN REF. 3; AA SEQUENCE).
 SQ SEQUENCE 785 AA; 80693 MW; 43893DBF6518B9EA CRC64;

Query Match 2.8%; Score 12; DB 1; Length 785;
 Best Local Similarity 100.0%; Pred. No. 0.0018;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 310 AHLRWHTGERPF 321
 |||||
 Db 645 AHLRWHTGERPF 656

RESULT 7
 SPI_RAT STANDARD; PRT; 788 AA.
 ID SPI_RAT
 AC Q01714;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Transcription factor Sp1.
 GN SPI.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=93010958; PubMed=1356762;
 RA Imataka H., Sogawa K., Yasumoto K., Kikuchi Y., Sasano K.;

Kobayashi A., Hayami M., Fujii-Kuriyama Y.;
 "Two regulatory proteins that bind to the basic transcription element
 gene.", a GC box sequence in the promoter region of the rat P-450IA1
 gene.", 11:3663-3671(1992).
 CC -1- FUNCTION: BINDS TO GC BOX PROMOTERS ELEMENTS AND SELECTIVELY
 CC ACTIVATES MRNA SYNTHESIS FROM GENES THAT CONTAIN FUNCTIONAL
 CC RECOGNITION SITES.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- PM: O-GLYCOSYLATED; CONTAINS N-ACETYLGLUCOSAMINE SIDE CHAINS
 CC (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE SP1 FAMILY OF C2H2-TYPE ZINC-FINGER
 CC PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D12768; BAA02235.1;
 CC PIR; JS0747; JS0747.
 CC HSSP; P08047; 1SP1.
 CC TRANSPAC; T00754;
 CC InterPro; IPR000822; Znf_C2H2.
 CC Pfam; PF00096; Zf-C2H2; 3.
 CC PRINTS; PR00048; ZINC_FINGER.
 CC ProDom; PD000003; Znf_C2H2; 2.
 CC SMART; SM00355; Znf_C2H2; 3.
 CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
 CC PROSITE; PS00157; ZINC_FINGER_C2H2_2; 3.
 CC Transcription regulation; Activator; Zinc-finger; Metal-binding;
 CC DNA-binding; Nuclear protein; Repeat; Glycoprotein.
 CC DOMAIN 629 711
 CC ZINC_FINGER.
 CC FT ZN_FING 629 711
 CC ZN_FING 659 683
 CC ZN_FING 689 711
 CC C2H2-TYPE.
 CC C2H2-TYPE.
 CC SEQUENCE 788 AA; 81015 MW; AA2B0CAB81AAB80C CRC64;
 CC -----
 CC Query Match 2.89; Score 12; DB 1; Length 788;
 CC Best Local Similarity 100.0%; Pred. No. 0.0018;
 CC Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC -----
 CC 310 AHLRWHTGTRPF 321
 CC |||||
 CC 648 AHLRWHTGTRPF 659
 CC -----
 CC RESULT 8
 CC BTD_DROME
 CC ID BTD_DROME STANDARD; PRT; 644 AA.
 CC AC Q24266; Q9W319;
 CC DT 15-JUL-1998 (Rel. 36, Created)
 CC DT 15-JUL-1998 (Rel. 36, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Transcription factor BTD (Butonhead protein).
 CC GN BTD OR CGI2653.
 CC OS Drosophila melanogaster (Fruit fly).
 CC OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 CC OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 CC OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 CC OX NCBI_TaxID=7227;
 CC [1]
 CC SEQUENCE FROM N.A.
 CC STRAIN-Canton-S; TISSUE-Embryo;
 CC MEDLINE-94081952; PubMed-8259212;
 CC Winner E.A., Jaekle H., Pfeifle C., Cohen S.M.;
 CC "A Drosophila homologue of human Sp1 is a head-specific segmentation
 CC gene.",
 CC Nature 366:690-694(1993).
 CC [2]

RP SEQUENCE FROM N.A.
 RC STRAIN-Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Crowley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo J.M., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Fertler S., Fleischmann W.,
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laspo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Fan S., Nusskern D.R., Pacleb J.M.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 CC -1- FUNCTION: REQUIRED FOR THE DEVELOPMENT OF THE ANTENNA,
 CC INTERCALARY AND MANDIBULAR SEGMENTS OF THE HEAD.
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -1- DEVELOPMENTAL STAGE: FIRST EXPRESSED IN A STRIPE COVERING THE HEAD
 CC ANLAGEN OF THE SYNCYTIAL BLASTODERM EMBRYO, PERSISTS THROUGH
 CC GASTRULATION AND DECAYS DURING GERM BAND EXTENSION, EXPRESSED
 CC LATER IN DEVELOPMENT IN A COMPLEX SPATIALLY RESTRICTED PATTERN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z29361; CAA82545.1;
 CC EMBL; AE003448; AAF46518.1;
 CC HSSP; P08047; 1SP2
 CC FlyBase; FBgn0000233; btd.
 CC InterPro; IPR000822; Znf_C2H2.
 CC Pfam; PF00096; Zf-C2H2; 3.
 CC PRINTS; PR00048; ZINC_FINGER.
 CC ProDom; PD000003; Znf_C2H2; 1.
 CC SMART; SM00355; Znf_C2H2; 3.
 CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
 CC PROSITE; PS00157; ZINC_FINGER_C2H2_2; 3.
 CC Transcription regulation; Activator; Zinc-finger; Metal-binding;
 CC DNA-binding; Nuclear protein; Repeat.
 CC DOMAIN 333 413
 CC ZINC_FINGERS.
 CC FT ZN_FING 333 357
 CC ZN_FING 363 385
 CC C2H2-TYPE.
 CC C2H2-TYPE.

```

FT ZN_FING 391 413 C2H2-TYPE.
FT DOMAIN 14 89 GLN-RICH.
FT DOMAIN 208 220 POLY-ALA.
FT DOMAIN 431 434 POLY-ALA.
FT DOMAIN 486 492 POLY-PRO.
FT DOMAIN 499 502 POLY-THR.
FT DOMAIN 515 519 POLY-SER.
FT DOMAIN 530 536 POLY-SER.
FT DOMAIN 596 599 POLY-SER.
SQ SEQUENCE 644 AA; 68581 MW; A0DB98C2AF938452 CRC64;

Query Match
Best Local Similarity 100.08; Pred. No. 0.014; Length 644;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 311 HLRWHTGERPF 321
Db 353 HLRWHTGERPF 363
|||||

RESULT 9
2014_XENLA STANDARD; PRT; 139 AA.
P18740:
01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DE 01-FEB-1994 (Rel. 28, Last annotation update)
DE Oocyte zinc finger protein XLCOF14 (Fragment).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90040698; PubMed=2509712;
RA Nietfeld W., El-Baradi T., Mentzel H., Pieler T., Koester M.,
RA Poeting A., Knoechel W.,
RT "Second-order repeats in Xenopus laevis finger proteins.";
RL J. Mol. Biol. 208:639-659(1989).
DR PIR: S06550; S06550.
DR HSP: P25490; 12NM.
DR InterPro: IPR001230; Ptenvyl.site.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; Zf-C2H2; 5.
DR ProDom: PD000003; Znf_C2H2; 1.
DR SMART: SM00355; Znf_C2H2; 5.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 5.
KW Zinc-finger; Metal-binding; DNA-binding; Repeat.
FT NON_TER 1 1
ZN_FING 6 28 C2H2-TYPE.
ZN_FING 33 55 C2H2-TYPE.
ZN_FING 61 83 C2H2-TYPE.
FT ZN_FING 89 111 C2H2-TYPE.
FT ZN_FING 117 139 C2H2-TYPE.
FT NON_TER 139 139
SQ SEQUENCE 139 AA; 15975 MW; E2B506C24185F4F2 CRC64;

Query Match
Best Local Similarity 100.08; Pred. No. 0.31; Length 139;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 315 HTGERPFVC 323
Db 111 HTGERPFVC 119
|||||

RESULT 10
BTE4_MOUSE STANDARD; PRT; 251 AA.
AC P58334;
DT 15-JUN-2002 (Rel. 41, Created)

```

```

DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcription factor BTEB4 (Basic transcription element binding-
DE protein 4) (BTE-binding protein 4) (Krueppel-like factor 16) (Dopamine
DE receptor regulating factor).
GN KLF16 OR BTEB4 OR DRRF.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21309923; PubMed=11390978;
RA Huang C.K., D'Souza U.M., Eisch A.J., Yajima S., Lammers C.-H.,
RA Yang Y., Lee S.-H., Kim Y.-M., Nestler E.J., Mouradian M.M.;
RT "Dopamine receptor regulating factor, DRRF: a zinc finger
RT transcription factor.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7558-7563(2001).
CC -!- FUNCTION: Transcription factor that binds GC and GT boxes in the
CC DIA, D2 and D3 dopamine receptor promoters and displaces Sp1 and
CC Sp3 from these sequences. It modulates dopaminergic transmission
CC in the brain by repressing or activating transcription from
CC several different promoters depending on cellular context.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: High expression in brain; olfactory tubercle,
CC olfactory bulb, nucleus accumbens, striatum, hippocampal CA1
CC region, amygdala, dentate gyrus and frontal cortex. Moderate
CC expression in hippocampal CA2-3 regions, piriform cortex, septum,
CC and distinct thalamic nuclei. Low expression in the cerebellum.
CC -!- DOMAIN: The Ala/Pro-rich domain may contain discrete activation
CC and repression subdomains and also can mediate protein-protein
CC interactions.
CC -!- SIMILARITY: BELONGS TO THE SP1 FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; AF283891; AAK66968.1;
CC MGD; MGI:2153049; Klf16.
CC TRANSFAC; T05053;
CC InterPro: IPR000822; Znf_C2H2.
CC Pfam; PF00096; Zf-C2H2; 3.
CC SMART; SM00355; Znf_C2H2; 3.
CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
CC PROSITE; PS00157; ZINC_FINGER_C2H2_2; 3.
KW Transcription regulation; DNA-binding; Nuclear protein; Repeat;
KW Zinc-finger; Metal-binding.
FT DOMAIN 3 136 ALA/PRO-RICH.
FT DOMAIN 103 116 SER-RICH.
FT DOMAIN 126 208 ZINC_FINGERS.
FT ZN_FING 126 150 C2H2-TYPE.
FT ZN_FING 156 180 C2H2-TYPE.
FT ZN_FING 186 208 C2H2-TYPE.
FT DOMAIN 223 248 PRO/SER-RICH.
SQ SEQUENCE 251 AA; 25665 MW; 3F0D773B1A09FA4 CRC64;

Query Match
Best Local Similarity 100.08; Pred. No. 0.52; Length 251;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 358 KRFRSDHL 366
Db 193 KRFRSDHL 201
|||||

RESULT 11

```

2134_HUMAN
ID P134_HUMAN STANDARD; PRT; 348 AA.
AC P52741;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc finger protein 134.
GN ZNF134.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Insulinoma; PubMed=7557990;
RX MEDLINE=96044430; PubMed=7557990;
TX Tomerup N., Vissing H.;
"Isolation and fine mapping of 16 novel human zinc finger-encoding
cDNAs identify putative candidate genes for developmental and
malignant disorders";
Genomics 27:259-264 (1995).
RL Genomics 27:259-264 (1995).
CC -!- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
FINGER PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U09412; AAC50253.1; -
DR HSSP; P08047; 1SP2.
DR Genew; HGNC:12918; ZNF134.
DR MIM; 604076; -
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; zf_C2H2; 9.
DR PRINTS; PR00048; ZINCFINGER.
DR PRODOM; PD000003; Znf_C2H2; 9.
DR SMART; SM00355; Znf_C2H2; 9.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 9.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 9.
DR Transcription regulation; Zinc-finger; DNA-binding; Metal-binding;
Nuclear protein; Repeat.
FT DOMAIN 97 343 ZINC FINGERS.
FT ZN_FING 97 119 C2H2-TYPE.
FT ZN_FING 125 147 C2H2-TYPE.
FT ZN_FING 153 175 C2H2-TYPE.
FT ZN_FING 181 203 C2H2-TYPE.
FT ZN_FING 209 231 C2H2-TYPE.
FT ZN_FING 237 259 C2H2-TYPE.
FT ZN_FING 265 287 C2H2-TYPE.
FT ZN_FING 293 315 C2H2-TYPE.
FT ZN_FING 321 343 C2H2-TYPE.
SQ SEQUENCE 348 AA; 40297 MW; E0941AD33EC8670 CRC64;
Query Match 2.18; Score 9; DB 1; Length 348;
Best Local Similarity 100.0%; Pred. No. 0.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 315 HTGERPFVC 323
DB 287 HTGERPFVC 295
RESULT 12
TF3A_HUMAN
ID TF3A_HUMAN STANDARD; PRT; 423 AA.
AC Q92664; Q13097; Q12963;
DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcription factor IIIA (Factor A) (TFIIIA).
GN GTF3A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Petal brain;
RX MEDLINE=95309028; PubMed=7789179;
RA Arakawa H., Nagase H., Hayashi N., Ogawa M., Nagata M.,
Fujiwara T., Takahashi E., Shin S., Nakamura Y.;
"Molecular cloning, characterization, and chromosomal mapping of a
novel human gene (GTF3A) that is highly homologous to Xenopus
transcription factor IIIA";
Cytogenet. Cell Genet. 70:235-238 (1995).
RN [2]
RP SEQUENCE OF 61-423 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95347600; PubMed=7622052;
RA Drew P.D., Nagle J.W., Canning R.D., Ozato K., Biddison W.E.,
Becker K.G.;
"Cloning and expression analysis of a human cDNA homologous to
Xenopus TFIIIA";
Gene 159:215-218 (1995).
RN [3]
RN CHARACTERIZATION.
RP MEDLINE=94342241; PubMed=8063702;
RA Moorefield B., Roeder R.G.;
"Purification and characterization of human transcription factor
IIIA";
J. Biol. Chem. 269:20857-20865 (1994).
CC -!- FUNCTION: INTERACTS WITH THE INTERNAL CONTROL REGION (ICR) OF
CC APPROXIMATELY 50 BASES WITHIN THE 5S RNA GENES. IS REQUIRED FOR
CC CORRECT TRANSCRIPTION OF THESE GENES BY RNA POLYMERASE III. ALSO
CC BINDS THE TRANSCRIBED 5S RNA. S. MAY INITIATE TRANSCRIPTION OF THE
CC 5S RIBOSOMAL RNA GENE AND MAINTAIN THE STABILITY OF TRANSCRIPTION
CC OF OTHER GENES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: UBQUITOUS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D32257; BAA06988.1; -
DR EMBL; U20272; AAA75623.1; -
DR EMBL; U14134; AAA21873.1; -
DR HSSP; P03001; 1TF3
DR TRANSFAC; T04953;
DR Genew; HGNC:4662; GTF3A.
DR MIN; 600860; -
DR InterPro; IPR000822; Znf_C2H2;
DR Pfam; PF00096; zf_C2H2; 9.
DR PRINTS; PR00048; ZINCFINGER.
DR SMART; SM00355; Znf_C2H2; 8.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 7.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 8.
DR Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
RNA-binding; Repeat; Nuclear protein.
FT DOMAIN 98 359 ZINC FINGERS.
FT ZN_FING 98 122 C2H2-TYPE.
FT ZN_FING 128 152 C2H2-TYPE.
FT ZN_FING 158 183 C2H2-TYPE.
FT ZN_FING 190 212 C2H2-TYPE (ATYPICAL).
FT ZN_FING 220 244 C2H2-TYPE.
FT ZN_FING 247 271 C2H2-TYPE.


```

RX MEDLINE=94327649; PubMed=8051167;
RA Imataka H., Nakayama K., Yasumoto K., Mizuno A., Fujii-Kuriyama Y.,
RA Hayami M.;
RT "Cell-specific translational control of transcription factor BTEB
RT expression. The role of an upstream AUG in the 5'-untranslated
RT region.";
RL J. Biol. Chem. 269:20668-20673(1994).
CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS TO GC BOX PROMOTER
CC ELEMENTS. SELECTIVELY ACTIVATES MRNA SYNTHESIS FROM GENES
CC CONTAINING TANDEM REPEATS OF GC BOXES BUT REPRESSES GENES WITH
CC A SINGLE GC BOX.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D31716; BAA06524.1; -
CC EMBL; S72504; AAD14110.1; -
CC HSSP; P08047; 1SP2.
CC TRANSFAC; T02212; -
CC Genew; HGNC:1123; BTEB1.
CC MIM; 602902; -
CC InterPro; IPR000822; Znf_C2H2.
CC Pfam; PF00096; zf-C2H2; 3.
CC PRINTS; PR00048; ZINCFINGER.
CC ProDom; PD000003; Znf_C2H2; 1.
CC SMART; SM00355; Znf_C2H2; 3.
CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
CC PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.
CC Transcription regulation; DNA-binding; Nuclear protein; Repeat;
KW Zinc-finger; Metal-binding.
FT DOMAIN 84 116 ASP/GLU-RICH (ACIDIC).
FT ZN_FING 143 225 ZINC FINGERS.
FT ZN_FING 173 197 C2H2-TYPE.
FT ZN_FING 203 225 C2H2-TYPE.
SQ SEQUENCE 244 AA; 27234 MW; 2D1B5A5BB9D42221 CRC64;

Query Match 1.9%; Score 8; DB 1; Length 244;
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 297 GCGKVGK 304
DB 149 GCGKVGK 156
|||||

RESULT 15
BTEB_MOUSE
ID BTEB_MOUSE STANDARD; PRT; 244 AA.
AC Q35739.
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcription factor BTEB1 (Basic transcription element binding
DE protein 1) (BTE-binding protein 1) (GC box binding protein 1)
DE (Kruppel-like factor 9).
GN BTEB1 OR BTEB-1 OR BTEB OR KLF9.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALE/C; TISSUE=Brain;
RX MEDLINE=99077958; PubMed=9858544;
RA Imhof A., Schuierer M., Werner O., Moser M., Roth C., Bauer R.,
RA Buettner R.;
```

```

RT "Transcriptional regulation of the AP-2alpha promoter by BTEB-1 and
RT AP-2rep, a novel wt-1/egr-related zinc finger repressor.";
RL Mol. Cell. Biol. 19:194-204(1999).
CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS TO GC BOX PROMOTER
CC ELEMENTS. SELECTIVELY ACTIVATES MRNA SYNTHESIS FROM GENES
CC CONTAINING TANDEM REPEATS OF GC BOXES BUT REPRESSES GENES WITH
CC A SINGLE GC BOX.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Y14296; CAA74671.1; -
CC HSSP; P08047; 1SP2.
CC TRANSFAC; T04700; -
CC MGD; MGI:1333856; Klf9.
CC InterPro; IPR000822; Znf_C2H2.
CC Pfam; PF00096; zf-C2H2; 3.
CC PRINTS; PR00048; ZINCFINGER.
CC ProDom; PD000003; Znf_C2H2; 1.
CC SMART; SM00355; Znf_C2H2; 3.
CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
CC PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.
CC Transcription regulation; DNA-binding; Nuclear protein; Repeat;
KW Zinc-finger; Metal-binding.
FT DOMAIN 84 116 ASP/GLU-RICH (ACIDIC).
FT ZN_FING 143 225 ZINC FINGERS.
FT ZN_FING 173 197 C2H2-TYPE.
FT ZN_FING 203 225 C2H2-TYPE.
SQ SEQUENCE 244 AA; 27169 MW; BBDF607FFA218D5A CRC64;

Query Match 1.9%; Score 8; DB 1; Length 244;
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 297 GCGKVGK 304
DB 149 GCGKVGK 156
|||||

Search completed: February 19, 2003, 13:38:02
Job time : 15 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 19, 2003, 13:35:37 ; Search time 94 Seconds
(without alignments)
938.172 Million cell updates/sec

Title: US-09-734-329-2
Perfect score: 428
Sequence: 1 MASSLLEEEAHYSSPLAML.....PAPPERAHGSSPEQSNLLEI 428

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_phc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	428	100.0	428	11	Q8VI67
2	127	29.7	431	4	Q8TDD2
3	33	7.7	452	4	Q96WJ1
4	33	7.7	523	5	Q9UIK4
5	33	7.7	666	5	Q9W3I8
6	25	5.8	398	11	Q9JHX2
7	14	3.3	367	13	Q90XF6
8	14	3.3	968	5	Q9VCB2
9	12	2.8	237	6	P79289
10	12	2.8	373	13	Q8UUU3
11	12	2.8	467	11	Q64I67
12	12	2.8	479	4	Q8WUW2
13	12	2.8	496	4	Q8WUW3
14	12	2.8	605	11	Q9D2H6
15	12	2.8	624	13	Q90WU2
16	12	2.8	713	4	Q8WUW4

17	12	2.8	725	11	O70494
18	12	2.8	760	13	Q90WR9
19	12	2.8	771	13	Q90WR8
20	12	2.8	781	11	O89090
21	12	2.8	782	11	O62445
22	12	2.8	783	11	O63158
23	12	2.8	784	11	O89087
24	10	2.3	278	13	Q8UUU4
25	10	2.3	298	5	Q93727
26	10	2.3	397	5	Q9W4F9
27	10	2.3	1263	5	P91639
28	10	2.3	1267	5	Q9VXH3
29	9	2.1	194	5	O62425
30	9	2.1	207	3	Q9UTAI
31	9	2.1	261	5	Q95U62
32	32	2.1	266	11	O9CSW1
33	34	2.1	341	5	Q8T362
34	9	2.1	387	11	Q9CSH8
35	9	2.1	400	11	Q8VHT7
36	9	2.1	411	13	P70043
37	9	2.1	411	13	P70046
38	9	2.1	419	5	O9N3C6
39	9	2.1	427	4	O9Y4B2
40	9	2.1	442	5	Q9VSF0
41	9	2.1	485	4	Q9NRY0
42	9	2.1	502	11	Q8R5B3
43	9	2.1	522	3	Q9HFF3
44	9	2.1	528	11	Q8R2M7
45	9	2.1	573	13	Q9PST7

ALIGNMENTS

RESULT 1

Q8VI67 Q8VI67 PRELIMINARY; PRT; 428 AA.
AC Q8VI67;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Osterix.
GN OSX OR C22.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21652532; PubMed=11792318;
RA Nakashima K., Zhou X., Kunkel G., Zhang Z., Deng J.M., Behringer R.R.,
de Crombrughe B.;
RT "The Novel Zinc Finger-Containing Transcription Factor Osterix is
Required for Osteoblast Differentiation and Bone Formation.";
RL Cell 108:17-29(2002).
DR EMBL; AF184902; AAL60067.1; .
DR MGD; MGI:2159409; Osx.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF000096; zf-C2H2; 3.
DR PRINTS; PR00048; ZINCFINGER.
DR PRODOM; PD000003; Znf_C2H2; 1.
DR SMART; SM00355; Znf_C2H2; 3.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_3.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.
SQ SEQUENCE 428 AA; 44718 MW; B79498958743586 CRC64;

Query Match 100.0%; Score 428; DB 11; Length 428;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 428; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MASSLLEEEAHYSSPLAMLTAACSKFGSGSPLRDSSTTLGKGTKKPYADLSAPKTMGDA 60
|||||
Db 1 MASSLLEEEAHYSSPLAMLTAACSKFGSGSPLRDSSTTLGKGTKKPYADLSAPKTMGDA 60

```
QY 61 YPAPSSNTGLSPAGSPAPASGAYANDYPPPHSPGPTGAQDPGLLVKPGHSSDCLP 120
|||||
DB 61 YPAPSSNTGLSPAGSPAPASGAYANDYPPPHSPGPTGAQDPGLLVKPGHSSDCLP 120
|||||
QY 121 SVYTSIDMTHPYGSYKAGIHAGISPGGNTPTPMDMHPGNGWLGOGGOGDGLQGTLS 180
|||||
DB 121 SVYTSIDMTHPYGSYKAGIHAGISPGGNTPTPMDMHPGNGWLGOGGOGDGLQGTLS 180
|||||
QY 181 TGAQPPPLNPQPTPSDFAPLNAPYAPAPHLQPGQHVLPQDQVYKPKAVGNSQGLEGS 240
|||||
DB 181 TGAQPPPLNPQPTPSDFAPLNAPYAPAPHLQPGQHVLPQDQVYKPKAVGNSQGLEGS 240
|||||
QY 241 GAAKPPRGAGTGGSGYAGSAGRSTCDPCNQELERLGAAGAAGLRKPIHISCHIPGCGK 300
|||||
DB 241 GAAKPPRGAGTGGSGYAGSAGRSTCDPCNQELERLGAAGAAGLRKPIHISCHIPGCGK 300
|||||
QY 301 VYKASHLKAHLRWHITGERPFCVNLFCGKRFTSRDELERHVRTTTRKKFTCLLCSKRF 360
|||||
DB 301 VYKASHLKAHLRWHITGERPFCVNLFCGKRFTSRDELERHVRTTTRKKFTCLLCSKRF 360
|||||
QY 361 TRSDHLSKHQTHRGEPGPPSPGKELGEGRSVGEENANPPRSTSPAPPEKAHGSP 420
|||||
DB 361 TRSDHLSKHQTHRGEPGPPSPGKELGEGRSVGEENANPPRSTSPAPPEKAHGSP 420
|||||
QY 421 EQSNLLEI 428
|||||
DB 421 EQSNLLEI 428
|||||

RESULT 2
Q8TDD2 PRELIMINARY; PRT: 431 AA.
AC Q8TDD2:
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Osterix.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP TISSUE=OSTEOSARCOMA;
RA Ganss B.W.;
RT "CDNA sequence, gene structure and chromosomal localization of the
RT human osterix (OSX) gene.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF477981; XAL84281.1;
SQ SEQUENCE 431 AA; 44994 MW; 454A6FEA84309FF9 CRC64;

Query Match 29.7%; Score 127; DB 4; Length 431;
Best Local Similarity 100.0%; Pred. No. 4.4e-118;
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 267 CDCPNCQELERLGAAGAAGLRKPIHISCHIPGCGKVYKASHLKAHLRWHITGERPFCVNL 326
|||||
DB 270 CDCPNCQELERLGAAGAAGLRKPIHISCHIPGCGKVYKASHLKAHLRWHITGERPFCVNL 329
|||||
QY 327 FCCKRFTSRDELERHVRTTTRKKFTCLLCSKRFTRSDHLSKHQTHRGEPGPPSPGPK 386
|||||
DB 330 FCCKRFTSRDELERHVRTTTRKKFTCLLCSKRFTRSDHLSKHQTHRGEPGPPSPGPK 389
|||||
QY 387 ELGEGRS 393
|||||
DB 390 ELGEGRS 396
|||||

RESULT 3
Q96MJ1 PRELIMINARY; PRT: 452 AA.
ID Q96MJ1
AC Q96MJ1;
```

```
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE CDNA FLJ2295 fis, clone PROST2001823, weakly similar to transcription
DE factor Sp1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PROSTATE;
RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
RA Houta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J., Chiba Y.,
RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,
RA Yanamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK056857; BAB71297.1;
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF000096; zf-C2H2_3.
DR PRODOM: PD000003; Znf_C2H2; 2.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_3.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 3.
KW DNA-binding; Zinc-finger.
SQ SEQUENCE 452 AA; 45836 MW; 95383C60C112320F CRC64;

Query Match 7.7%; Score 33; DB 4; Length 452;
Best Local Similarity 100.0%; Pred. No. 3.9e-24;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 306 SHLKAHLRWHITGERPFCVNLFCGKRFTSRDEL 338
|||||
DB 333 SHLKAHLRWHITGERPFCVNLFCGKRFTSRDEL 365
|||||

RESULT 4
Q9ULK4 PRELIMINARY; PRT: 523 AA.
ID Q9ULK4
AC Q9ULK4:
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE C2H2 zinc finger transcription factor.
GN Sp1 OR D-Sp1 OR CG1343.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Eurygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CANTON S;
RA MEDLINE=20025540; PubMed=10559487;
RA Schock F., Purnell B.A., Wimmer E.A., Jackle H.;
RT "Common and diverged functions of the Drosophila gene pair D-Sp1 and
RT buttonhead.";
RL Mech. Dev. 89:125-132(1999).
DR EMBL: AJ131022; CAB55429.1;
DR HSSP: P08047; 1SP2.
DR FlyBase: FBgn0020378; Sp1.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF000096; zf-C2H2; 3.
DR PRODOM: PD000003; Znf_C2H2; 1.
DR SMART: SM00355; Znf_C2H2; 3.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 3.
KW DNA-binding; Metal-binding; Zinc-finger.
SQ SEQUENCE 523 AA; 54577 MW; 64FB8392A72A7A8B CRC64;
```

Query Match 7.7% Score 33; DB 5; Length 523;
Best Local Similarity 100.0%; Pred. No. 4.5e-24;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 306 SHLKAHLRWHGTGERPFCVNLFCGKRTSRDEL 338
|||||
DB 205 SHLKAHLRWHGTGERPFCVNLFCGKRTSRDEL 237
|||||

RESULT 5
Q9W318 PRELIMINARY; PRT; 666 AA.
AC Q9W318:
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
SPL protein.
SPL OR CG1343.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.A.,
RA Brandon R.C., Rogers J.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balleson R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Beriman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Fertaz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.B., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleib J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RL "The genome sequence of Drosophila melanogaster";
Science 287:2185-2195(2000).
EMBL: AE003448; AAF46519.1;
DR HSP; P08047; ISP2
DR FlyBase; FBgn020378; Spl.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; Zf-C2H2; 3.
DR ProDom; PD000003; Znf_C2H2; 2.
DR SMART; SM00355; Znf_C2H2; 3.

DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 3.
DR DNA-binding; Metal-binding; zinc-finger.
SQ SEQUENCE 666 AA; 67658 MW; 9D4AF2B1756D1148 CRC64;

Query Match 7.7% Score 33; DB 5; Length 666;
Best Local Similarity 100.0%; Pred. No. 5.6e-24;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 306 SHLKAHLRWHGTGERPFCVNLFCGKRTSRDEL 338
|||||
DB 351 SHLKAHLRWHGTGERPFCVNLFCGKRTSRDEL 383-
|||||

RESULT 6
Q9JHX2 PRELIMINARY; PRT; 398 AA.
AC Q9JHX2:
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Zinc finger protein Sp5.
GN SP5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6XCBa;
RX MEDLINE=20325586; PubMed=11071760;
RA Harrison S.M., Houzelstein D., Dunwoodie S.L., Baddington R.S.P.;
RT "Sp5, a new member of the Spl family, is dynamically expressed during
RL development and genetically interacts with Brachyury.";
Dev. Biol. 227:358-372(2000).
DR EMBL; AF279479; AAF87798.1;
DR HSP; P08047; ISP1.
DR MGD; MGI:1927715; Sp5.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; Zf-C2H2; 3.
DR ProDom; PD000003; Znf_C2H2; 2.
DR SMART; SM00355; Znf_C2H2; 3.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 3.
KW DNA-binding; Metal-binding; zinc-finger.
SQ SEQUENCE 398 AA; 42052 MW; B35985375586B5C9 CRC64;

Query Match 5.8% Score 25; DB 11; Length 398;
Best Local Similarity 100.0%; Pred. No. 3.5e-16;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 306 SHLKAHLRWHGTGERPFCVNLFCGK 330
|||||
DB 311 SHLKAHLRWHGTGERPFCVNLFCGK 335
|||||

RESULT 7
Q90XF6 PRELIMINARY; PRT; 367 AA.
ID Q90XF6:
AC Q90XF6:
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Zinc finger buttonhead-related transcription factor 1.
GN BTSL.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21521131; PubMed=11641225;

RA Tallafuss A., Wilm T.P., Crozatier M., Pfeffer P., Wassef M.,
 RA Bally-Cuif L.,
 RT "The zebrafish buttonhead-like factor Bts1 is an early regulator of
 RT pax2.1 expression during mid-hindbrain development.",
 RL Development 128:4021-4034(2001).
 RA EMBL: AF388363; AAK83353.1; -
 DR InterPro: IPR000822; Znf_C2H2.
 DR Pfam: PF00096; zf-C2H2_3.
 DR ProDom: PD000003; Znf_C2H2_1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_3.
 DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 3.
 KW DNA-binding; Zinc-finger.
 SQ SEQUENCE 367 AA; 41030 MW; F31946B6A66CD8FD CRC64;

Query Match 3.3%; Score 14; DB 13; Length 367;
 Best Local Similarity 100.0%; Pred. No. 3.1e-05;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 317 GERPFVCNWLFCGK 330
 DB 288 GERPFVCNWLFCGK 301
 ||||||||||||||||
 ||||||||||||||||

ULT 8
 -VCB2
 ID Q9VCB2 PRELIMINARY; PRT; 968 AA.
 AC Q9VCB2;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DE CG5669 protein (LD04007p)..
 GN CG5669
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brattier P.,
 RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Foster K., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jaiswal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle B.J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.-H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
 RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
 RA Miranda A., Mungall C.J., Nunoo J., Pacle B.J., Paragas V., Park S.,
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
 RA Celniker S.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AEO03746; AAF56261.1; -
 DR EMBL: AY089533; AAL90271.1; -
 DR HSSP: P08047; 1SP2.
 DR FlyBase: FBgn0039169; CG5669.
 DR InterPro: IPR000822; Znf_C2H2.
 DR Pfam: PF00096; zf-C2H2_3.
 DR ProDom: PD000003; Znf_C2H2_2.
 DR SMART: SM00355; Znf_C2H2_3.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 3.
 DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 3.
 KW DNA-binding; Metal-binding; Zinc-finger.
 SQ SEQUENCE 968 AA; 102020 MW; 834CB19340C302CB CRC64;

Query Match 3.3%; Score 14; DB 5; Length 968;
 Best Local Similarity 100.0%; Pred. No. 7.7e-05;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 310 AHLRWHTGERPFVC 323
 DB 759 AHLRWHTGERPFVC 772
 ||||||||||||||||
 ||||||||||||||||

RESULT 9
 P79289
 ID P79289 PRELIMINARY; PRT; 237 AA.
 AC P79289;
 DT 01-MAY-1997 (TREMBLrel. 03, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Spt transcription factor (Fragment).
 OS Spt scrofa (Big).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-ENDOMETRIUM;
 RX MEDLINE=20150569; PubMed=10687861;
 RA Simmen R.C.M., Zhang X.L., Zhang D., Wang Y., Michel F.J.,
 RA Simmen F.A.;
 RT "Expression and regulatory function of the transcription factor Spt in
 RT the uterine endometrium at early pregnancy: implications for
 RT epithelial phenotype";
 RL Mol. Cell. Endocrinol. 159:159-170(2000).
 DR EMBL: U57347; AAB39513.3; -
 DR HSSP: P08047; 1SP2.
 DR InterPro: IPR000822; Znf_C2H2.
 DR Pfam: PF00096; zf-C2H2_3.
 DR ProDom: PD000003; Znf_C2H2_2.
 DR SMART: SM00355; Znf_C2H2_3.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 3.
 DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 4.
 KW DNA-binding; Metal-binding; Zinc-finger.
 FT NON TER 1
 FT NON TER 237
 SQ SEQUENCE 237 AA; 25421 MW; C6950DB42912DAB6 CRC64;

Query Match 2.8%; Score 12; DB 6; Length 237;
Best Local Similarity 100.0%; Pred. No. 0.0021;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 310 AHLRWHTGERPF 321
|||||
DB 115 AHLRWHTGERPF 126

RESULT 10
Q8UUU3 PRELIMINARY; PRT; 373 AA.
AC Q8UUU3

DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE SPl-like zinc-finger protein XSPR-1.
Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;

RN [1]
RP SEQUENCE FROM N.A.
RA Ossipova O., Stick R., Pieler T.;
RT "Interaction between two novel SPl-like zinc finger proteins and
Brachyury in xenopus.";

RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY062264; AAL47217.1;

DR InterPro; IPR000822; Znf_C2H2.

DR Pfam; PF00096; zf-C2H2; 3.

DR PRINTS; PR00048; ZINC_FINGER.

DR PRODOM; PD000003; Znf_C2H2; 1.

DR SMART; SM00355; Znf_C2H2; 3.

DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_3.

DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 3.

KW Zinc.

SQ SEQUENCE 373 AA; 41496 MW; EBF7A4F648153CDA CRC64;

Query Match 2.8%; Score 12; DB 13; Length 373;
Best Local Similarity 100.0%; Pred. No. 0.0032;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 CHIPCGKGVYCK 304
|||||
273 CHIPCGKGVYCK 284

RESULT 11

Q64167 PRELIMINARY; PRT; 467 AA.
AC Q64167; Q62251;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Trans-acting transcription factor 1 (Transcription factor SPl) (SPl
DE gene) (3' end).

GN SPl.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=96016118; PubMed=7568082;

RA Persengiev S.P., Saffer J.D., Kilpatrick D.L.;

RT "An alternatively spliced form of the transcription factor SPl
containing only a single glutamine-rich transactivation domain.";

RL Proc. Natl. Acad. Sci. U.S.A. 92:9107-9111(1995).

RN [2]

RP SEQUENCE OF 367-467 FROM N.A.

RX MEDLINE=92338398; PubMed=1633330;

RA Chestier A., Charnay P.;

RT "Difference in the genomic organizations of the related transcription
factors SPl and Krox-20; possible evolutionary significance.";
RL DNA Seq. 2:325-327(1992).

CC -(- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).

DR EMBL; S79832; AAB35321.1;

DR EMBL; X60136; CAA42721.1;

DR HSSP; P08047; ISPI.

DR MGD; MGI:98372; SPl.

DR InterPro; IPR000822; Znf_C2H2.

DR Pfam; PF00096; zf-C2H2; 3.

DR PRINTS; PR00048; ZINC_FINGER.

DR PRODOM; PD000003; Znf_C2H2; 2.

DR SMART; SM00355; Znf_C2H2; 3.

DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.

DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 3.

KW DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.

SQ SEQUENCE 467 AA; 48749 MW; 4A373D67E6128197 CRC64;

Query Match 2.8%; Score 12; DB 11; Length 467;
Best Local Similarity 100.0%; Pred. No. 0.0039;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 310 AHLRWHTGERPF 321
|||||

DB 329 AHLRWHTGERPF 340

RESULT 12

Q8WUW2 PRELIMINARY; PRT; 479 AA.

ID Q8WUW2

AC Q8WUW2

DT 01-MAR-2002 (TrEMBLrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Specificity protein 3 internally initiated isoform 2.

GN SP3.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Hernandez E.M., Johnson A., Notario V., Chen A., Richert J.R.;

RT "AVA as a Translation Initiation Site in Vitro for the Human

Transcription Factor Sp3.";

RL J. Biochem. Mol. Biol. 0:0-0(2002).

DR EMBL; AY070137; AAL58088.1;

DR InterPro; IPR000822; Znf_C2H2.

DR Pfam; PF00096; zf-C2H2; 3.

DR PRINTS; PR00048; ZINC_FINGER.

DR PRODOM; PD000003; Znf_C2H2; 1.

DR SMART; SM00355; Znf_C2H2; 3.

DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_3.

DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 3.

SQ SEQUENCE 479 AA; 52064 MW; DA04C6B5CBB307A CRC64;

Query Match 2.8%; Score 12; DB 4; Length 479;
Best Local Similarity 100.0%; Pred. No. 0.004;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 CHIPCGKGVYCK 304
|||||

DB 321 CHIPCGKGVYCK 332

RESULT 13

Q8WUW3 PRELIMINARY; PRT; 496 AA.

ID Q8WUW3

AC Q8WUW3

DT 01-MAR-2002 (TrEMBLrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Specificity protein 3 internally initiated isoform 1.

```
GN SP3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Hernandez E.M., Johnson A., Notario V., Chen A., Richert J.R.;
RT "AUA as a Translation Initiation Site in Vitro for the Human
RL Transcription Factor Sp3.";
RL J. Biochem. Mol. Biol. 0:0-0(2002).
DR EMBL; AY070137; AAL58087.1; -.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 3.
DR PRINTS; PRO0048; ZINC_FINGER.
DR ProDom; PD000003; Znf_C2H2; 1.
DR SMART; SM00355; Znf_C2H2; 3.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_3.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.
DR SEQUENCE 496 AA; 53729 MW; 65658A6B55548661 CRC64;
SQ
Query Match 2.8%; Score 12; DB 4; Length 496;
Best Local Similarity 100.0%; Pred. No. 0.0041;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 293 CHIPGCGKVGK 304
DB 338 CHIPGCGKVGK 349
IIIIIIIIIIII
RESULT 14
Q9D2H6 PRELIMINARY; PRT; 605 AA.
AC Q9D2H6;
CT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE 49304801168ik protein (Sp2 transcription factor).
GN SP2 OR 49304801168IK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RN Nature 409:685-690(2001).
RP [2]
RP SEQUENCE FROM N.A.
RC TISSUE=MAMMARY TUMOR;
RA Strausberg R.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK019649; BAB31823.1; -.
```

```
DR EMBL; BC021759; AAH21759.1; -.
DR HSPF; P08047; 1SP2.
DR MGD; MGI:1926162; Sp2.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 3.
DR ProDom; PD000003; Znf_C2H2; 2.
DR SMART; SM00355; Znf_C2H2; 3.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 2.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_2; 3.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.
DR DNA-binding; Metal-binding; Zinc-finger.
SQ SEQUENCE 605 AA; 64131 MW; 2B48ACE7A8EF1B70 CRC64;
Query Match 2.8%; Score 12; DB 11; Length 605;
Best Local Similarity 100.0%; Pred. No. 0.005;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 327 FCGKRFRSDEL 338
DB 553 FCGKRFRSDEL 564
IIIIIIIIIIII
RESULT 15
Q90WU2 PRELIMINARY; PRT; 624 AA.
ID Q90WU2;
AC Q90WU2;
CT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE SP3 transcription factor (Fragment).
OS Fundulus heteroclitus (Killifish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Cyprinodontiformes; Fundulidae; Fundulus.
OX NCBI_TaxID=8078;
RN [1]
RP SEQUENCE FROM N.A.
RA Kottell K.J., Crawford D.L.;
RT "Evolution of Sp Transcription Factors.";
RL Mol. Biol. Evol. 0:0-0(2001).
DR EMBL; AY057451; AAL23671.1; -.
DR InterPro; IPR000276; GPCR_Rhodopsn.
DR InterPro; IPR003880; Ppantme_attach.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 3.
DR ProDom; PD000003; Znf_C2H2; 1.
DR PROSITE; PS00237; G_PROTEIN_RECP_F1_1; UNKNOWN_1.
DR PROSITE; PS00012; PHOSPHOPANTHETINE; UNKNOWN_1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_3.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.
KW DNA-binding; Zinc-finger.
FT NON_TER
SQ SEQUENCE 624 AA; 65296 MW; B877AB4E4A4AC32F1 CRC64;
Query Match 2.8%; Score 12; DB 13; Length 624;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 293 CHIPGCGKVGK 304
DB 487 CHIPGCGKVGK 498
IIIIIIIIIIII
Search completed: February 19, 2003, 13:39:43
Job time : 95 secs
```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 19, 2003, 13:21:11 ; Search time 84 Seconds

(Without alignments)
678.944 Million cell updates/sec

Title: US-09-734-329-2

Perfect score: 2384

Sequence: 1 MASSLLLEEAHYGSSPLAML.....PAPPEKANGSPESQSNLLEI 428

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002:*
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	676	28.4	554	22 ABG15134	Novel human diagno
2	663.5	27.8	666	22 ABE58019	Drosophila melanog
3	544.5	22.8	398	23 AAM49117	Human Spl family t
4	543.5	22.8	398	23 AAM49116	Mouse Spl family t
5	471	19.8	785	22 AAM39393	Human polypeptide
6	471	19.8	794	22 AAM41179	Human polypeptide
7	461.5	19.4	968	22 ABE61691	Drosophila melanog
8	409	17.2	644	22 ABE70436	Drosophila melanog
9	406	17.0	612	23 ABG62110	Human prostate spe
10	401.5	16.8	135	23 ABP41177	Human ovarian anti

11	401	16.8	168	17 AAR95975	Nucleic acid recog
12	401	16.8	241	17 AAR96015	Target binding ass
13	401	16.8	273	17 AAR96011	Target binding ass
14	394.5	16.5	288	21 AAY95986	Human zinc finger
15	392	16.4	297	18 AAW08955	Chimeric restricti
16	391	16.4	94	21 AAB14293	Transcription fact
17	391	16.4	94	23 ABY03953	Zinc finger protei
18	391	16.4	94	23 ABP48202	Sp-1 zinc finger p
19	391	16.4	303	18 AAW08956	Chimeric restricti
20	377.5	15.8	169	22 AAM86021	Human immune/haema
21	377.5	15.8	169	22 AAU16476	Human novel secret
22	366.5	15.4	196	21 AAB07701	Zinc finger protei
23	366.5	15.4	196	22 AAE08714	Human ZFP-vascular
24	366.5	15.4	196	23 AAE21123	5 finger protein u
25	366.5	15.4	196	23 ABB07136	6-finger zing fing
26	359	15.1	99	21 AAB07699	Zinc finger protei
27	359	15.1	99	22 AAE08712	Human ZFP-vascular
28	359	15.1	99	23 AAE21124	3 finger protein u
29	359	15.1	99	23 ABB07131	Human veg 1 domain
30	353.5	14.8	470	20 AAY24314	Human repressor kr
31	353.5	14.8	470	21 AAB13773	Human RKLf protein
32	353.5	14.8	470	22 AAB48018	Human GKLf protein
33	353.5	14.8	479	20 AAY24315	Human repressor kr
34	353.5	14.8	479	21 AAB13774	Human RKLf protein
35	351	14.7	97	22 AAB62172	Zinc finger protei
36	351	14.7	99	21 AAB07700	Zinc finger protei
37	351	14.7	99	22 AAE08713	Human ZFP-vascular
38	351	14.7	99	23 ABB07135	Human Veg 3a domai
39	338.5	14.2	426	22 ABB60543	Drosophila melanog
40	337	14.1	469	22 AAB90780	Human shear stress
41	337	14.1	480	18 AAW19928	Transforming growt
42	336.5	14.1	474	23 ABB57314	Mouse ischaemic co
43	331	13.9	410	22 ABB64734	Drosophila melanog
44	320.5	13.4	480	23 ABB57129	Mouse ischaemic co
45	318	13.3	239	22 AAG75126	Human colon cancer

ALIGNMENTS

RESULT 1
ABG15134
ID ABG15134 standard; Protein; 554 AA.
XX AC ABG15134;
XX AC
DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #15125.
XX DE

Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX N-PSDB; AAS79321.

XX New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
PS Claim 20; SEQ ID NO 45493; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG0010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 554 AA;

Query Match 28.4%; Score 676; DB 22; Length 554;
Best Local Similarity 39.3%; Pred. No. 2.3e-41; Indels 94; Gaps 22;
Matches 171; Conservative 41; Mismatches 129;
QY 18 AMLTAACSKFGSSPLRSTTGKTKKPYADLSAPKTMGDYAPAPSSNGLLSP-AG 76
DB 76 AMLTAVCG-----SLGSOHTTEPHA--SPPL--DL--QPLQTYGHTSPGAG 117
QY 77 SPAPASGAYNDYPPPHSFP-GPTGAQDPGLLVKPHGSSDC-----LPSVYTSL 126
DB 118 DYPSPQLQ-----PGLQSLPLGPEVDFSGYELFGASSRVTCEDLESPLAPGPFSL 171
QY 127 ---DMTHPYGSKYKAGIAGISPGNGNTPTWDMHPCGNLGGGOGDGLGTLSTGP 183
DB 172 LQPDMSHYESWFRP-THPGAEDG-----SWNDLHPTGSWMDLPHTQG-----ALTSP 218
QY 184 AQP-PLNPQLPTPSDFAPLNPAPYPAPHLLQ--GPQHVLPQD-----VYKPKAYGN 233
DB 219 GHGPGALAGLGGYGDHQLCAPPPHAPHLLPRAAGGQHLGLPPDGAKALEVAAPESQGL 278
QY 234 SQLEGSGAARPPRGAGTGGGGYAGSAGRSTCDPCNQBELRLGAAA--GLRKKPIH 291
DB 279 DSSLQ--GAARPK-----GRRSVPRSSGQTVCRPCNLEAERLAPCGPDGKKKHLH 330
QY 292 SCHIPGCKGVYKASHLAHLRWHYGERPFVCNWLFCGKRTSRDELPRHVRHTREKKF 351
DB 331 NCHIFCGKAYAKTSHLAHLRWHSGDRFPVCNWLFCGKRTSRDELQRLQHTHTGKTF 390
QY 352 TLLCSKRTSRDHLRSKHQRTHEGPGPPGPKLEGRSVGEEAN---QPP-----403
DB 391 PCAVCSRVRMRSDHLAKHMKTH-----EGAKEERAAGAASGEGKAGAVEPPGGKX 441
QY 404 -----RSTSPAPP 412
DB 442 REAGSMASSPDSP 456
RESULT 2
ID ABB58019 standard; Protein: 666 AA.
XX
XX ABB58019;
XX
XX 26-MAR-2002 (first entry)
DT

XX Drosophila melanogaster polypeptide SEQ ID NO 849.
DE
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
XX 23-MAR-2000; 2000US-191637P.
PR
XX
XX 11-JUL-2000; 2000US-0614150.
PR
XX
XX (PEKE) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX
XX WPI; 2001-656860/75.
XX
XX N-PSDB; ABL02122.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Disclosure; SEQ ID NO 849; 21pp + Sequence Listing; English.
PS
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABBS7737-ABBS72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 666 AA;
SQ
Query Match 27.8%; Score 663.5; DB 22; Length 666;
Best Local Similarity 40.2%; Pred. No. 2.4e-40;
Matches 176; Conservative 37; Mismatches 138; Indels 87; Gaps 20;
QY 15 SPLAMLTAAACSKFGSSPLRSTTGKTKKPYADLSAPKTMGDYAPAPSS 67
DB 59 SPCAISAASSSSSGSGGQSSRLSSASTMVTNITASRPLA--SSCAAVGGGTGSSSS 116
QY 68 TNGLLSPAGSPAPASGAYNDYPPPHSFPPTGAQD-----PGLLVKPHGSSDCPLSPV 122
DB 117 ASGQS--SSTASAVAAAYGGDL-YFPNT--STSNMDNHHMQGLL-CKVAGAAAFGV 170
QY 123 YTSLDMTHPYGSKYKAGIH---AGISPGNGNTPTWDMH-PGNNLGGGOGDGLGTL 178
DB 171 YS---RHPYDWPENAVTHKEAASVNSG-----WDMHSAAGSMLDMG---GAGMST 216
QY 179 LSTGPAQPLNPQLPTPSDFAPLNPAPYPAPHLLQCPQHVLPQDVKPK-----229
DB 217 MA-----NYASENYSSALS-----HSLGSGQHL--QDTYKSLMPGQGVG 257
QY 230 -AVGNSG-----QLEGSGAARPPRGAGTGGG--GGYAGS-----AGSTCDPCNC 272
DB 258 VGVGNGFSLPHSPSFAAAAAATAAAGSGPQSGSSTPSRSQRYAGRATCCPNC 317
QY 273 QELERLGAAGAAAGLRKKPIHSCHIPGCKGVYKASHLAHLRWHYGERPFVCNWLFCGKRF 332
DB 318 QEAERLGPAGVHLRKKNIHSCHIPGCKGVYKASHLAHLRWHYGERPFVCNWLFCGKRF 377
QY 333 TRSELERHVRHTREKKFTCLCSKRTSRDHLRSKHQRTHEGPGPPGPKLEGRSVGEEAN---GEPGPGPPSGPKEL 388

DR WPI: 2002-114585/15.
DR N-PSDB; ABA96948.
XX
PT Spl family transcription factor having zinc finger domain for use as a
PT target for drug development
XX
PS Claim 1; Fig 1; 96pp; Japanese.
XX
CC This sequence represents a murine Spl family transcription factor,
CC m285, and its human homologue, h285 (AAM49117) contain 3
CC zinc finger domains and are present in the nucleus, where they bind GC
CC boxes and regulate transcriptional activity. The invention also
CC encompasses fragments of the 285 proteins and nucleic acids encoding
CC them, expression vectors and host cells comprising 285 protein-encoding
CC nucleic acids, the recombinant production of 285 proteins, antibodies
CC against 285 proteins, a method of screening for compounds which bind to
CC 285 proteins, and a method for the regulation of the transcription
CC activity of promoters, especially of the CMV promoter, using the 285
CC protein. The 285 proteins can be used target molecules for drug
CC development and for gene therapy.
XX
SQ Sequence 398 AA;
Query Match 22.8%; Score 543.5; DB 23; Length 398;
Best Local Similarity 34.1%; Pred. No. 8e-32;
Matches 142; Conservative 34; Mismatches 115; Indels 125; Gaps 16;
QY 15 SPLAMLTACSKFG--GSSPLRDLSTLKGKTKPY-----ADLSAPKTM 57
DB 31 SPLALLAATCSRIGQPGAAAPDFLQV-----PYDPALGSPSLRPHPTADM--PAHS 81
QY 58 GDAYAPSPSTNGLISPAAGSPAPASGYANDYPFPFPGTGAQDPLGLVPGHSS-- 115
DB 82 PGALPPPHPSLG--LTPQKTHLQPSFGAAHELPLTPADPSYPYEFSPVKMLPSMAALP 139
QY 116 SCLPSVSTLDMTHPGYSWKAGTHAG-----ISPQGNTPPTPW 156
DB 140 ASCAPAYV-----PYAA--QAALPPGYSNLLPPPPPPPTCRQLSPAPAPDPLPW 190
QY 157 DMHPPGNLGGQGGDGLQGT-LSTGPAQPLNPLQPTY-----PS 197
DB 191 SIPOS-----GAGPG--SSGVPGTSLSACAPPHAPRFPASAAAAAALQRLGVLGPS 245
QY 198 DFAPLNAPYAPHLLQPGQHVLPQDVYKPAVNSGQLESGAAGKPPRGAGTGGSGGY 257
DB 246 DFAQY-----QSCIAALQTKAP-----L 264
QY 258 AGSGAGRSTDCPCNQCQLERLGAAGAGLRKPIHSCHIPGGGKYGKASHLKAHLRWHTG 317
DB 265 AATARRCRRCPNCQAAG--CAPEAEPGKKQHVCHVPGCGKYGKTSHLKAHLRWHTG 322
QY 318 ERPFVNCNLFQCKRTRSDERHVRHTREKKFTCLCSKRFTSRDLSKHQTH 373
DB 323 ERPFVNCNLFQCKRTRSDERHVRHTREKKFTCLCSKRFTSRDLSKHQTH 378
RESULT 5
AAM39393
ID AAM39393 standard; Protein; 785 AA.
XX
AC AAM39393;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 2538.
XX
KW Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX

OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-052317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
DR N-PSDB; AAI58549.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 4; SEQ ID NO 2538; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nontropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 785 AA;
Query Match 19.8%; Score 471; DB 22; Length 785;
Best Local Similarity 45.0%; Pred. No. 3.5e-26;
Matches 107; Conservative 15; Mismatches 78; Indels 38; Gaps 6;
QY 177 GTLSTGPAQPLNPOLPT-----YPSDFAPLNAPYAPHLLQPGQHVLPQ 223
DB 514 GTVTVNAQAQLSSMPGLQTNLSALGTSIQVHPITQGLPLATIANAPGDHGAQLGHAGD 573
QY 224 DVYKPAVNSGQLESGAAGKPPRGAGTGGSGYAGSGAGRSTDCPCNQCQLERLGAAGA 283
DB 574 GIHDDTAGEEG--ENSPDAQP-----QAGRTRREACTCPYCKDESGRSGDP 620
QY 284 GLRKKPIHSCHIPGGKYGKASHLKAHLRWHTGPRFVNCNLFQCKRFTSRDELRHVR 343
DB 621 G--KKQKHICHQCGKYGKTSHLRAHLRWHTGPRFVNCNLFQCKRFTSRDELRHVR 678
QY 344 THTREKKFTCLCSKRFTSRDLSKHQTH--GEPPG-----PPSGPKELGEG 391
DB 679 THTREKKFTCLCSKRFTSRDLSKHQTH--GEPPG-----PPSGPKELGEG 736
RESULT 6
AAM41179


```
Db 494 GLGQVQIHANQLPPNLPANF-----QVLTQLKPSHP----- 526
QY 136 YKAGIHAGISPGNGTPTPWMDHPCGNWLGCGGGGGLQGTSTGPAQPLNP-QLPT 194
Db 527 -----QVQ---TQGVQVMPKQEPQSPQMIT 550
QY 195 Y-----PSDFAPLN-----PAPYAPHLLOPGPHV-----L 221
Db 551 SKQEPDPTFGPISATGNPPAPASTNTASPOOQIKFLHTESNLSLSIPASTQITAL 610
QY 222 PQDV-----YKPAVGN-----GOL-----EGSGAAKPP 246
Db 611 PQQATNTPTATPIVSLPARSKVNAVTTSSQITTIATPGGQVSVTTQARGATASI 570
QY 247 R-----GAGTGGGGYAGS-----GAGRSTCDPCNQB 275
Db 671 RSTNTSTTTTTPSQSHLNNNISVASGGAATGGGGGTATGPKPLKRVACTCPNCTDG 730
QY 276 ERLGAAAGLRKKPIHSCHPGCGKYGKASHLKAHLRWHHTGERFVGNWLCGRFTRS 335
Db 731 EKHS-----DKRQHICHTGCHKYVYKTSHLRAHLRWHHTGERFVCSWAFCGRFRS 784
336 DELERHVRTHREKFKETCLCSKRFTRSDHLSKHORTHCEPGPGPPPKELGERSVG 395
Db 785 DELQRHRTHTGKRCQCECNKFMRSKHLKHIKTHFKRSRG-----VELIELSIKQ 838
QY 396 EEEANOPPRS 405
Db 839 ETKGNAPKS 848

RESULT 8
ID ABB70436 standard; Protein: 644 AA.
XX
AC ABB70436;
DT
XX
26-MAR-2002 (first entry)
DE
XX
Drosophila melanogaster polypeptide SEQ ID NO 38100.
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
23-MAR-2001; 2001WO-US09231.
PR
XX
23-MAR-2000; 2000US-191637P.
XX
11-JUL-2000; 2000US-0614150.
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
DR N-PSDB; ABL14539.
PT
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
XX Disclosure; SEQ ID NO 38100; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
```

```
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLI16176-ABLI30511), expressed DNA
CC sequences (ABLI01840-ABLI16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 644 AA;
Query Match 17.2%; Score 409; DB 22; Length 644;
Best Local Similarity 28.8%; Pred No. 9.7e-22;
Matches 132; Conservative 34; Mismatches 122; Indels 170; Gaps 18;
QY 65 FSSTNGLISP-----AGSPAPASGYANDY--PPFPHSPFGP-----TGAODPGLLVKPGHS 114
Db 92 FLSSAALLSAPPSLSGSSGSSGSSPLYPKPKMLKLELPYQASSTGTASPNSSIQSAPS 151
QY 115 SSDCLPSVYVTSMDWTHPYGWSYKAGIHAGISPGNGTPTPWMDHPCGNWLGCGGGG 174
Db 152 SASVSPSIFPS-----PAQSFASIASP-STPTT----- 179
QY 175 LQGTSTGPAOPLNP---QLPTYPSDFAPLNPAPYPAPHLLOPGPHVLPQDVYKPAV 231
Db 180 ---TLA-----PPTTAAGALAGSTSSPSSSAASAA-----AAAAAA 216
QY 232 GNSGQLESGAAKPPRGAGTGGSG-----GYAGSGAGRST----- 266
Db 217 AAAADLGAASAAVAGWNTAYSGLGPAPRSQFPYQAYASDYGNVAGMSSAAWFSHOER 276
QY 267 -----CDCPNC-QELERL-----GAAAAGLRKKPIH 291
Db 277 LYQPWSSSYPGNFDDIAFQLOLRSSVRCCTPCNCTNEMSGLPPIVGPDERG-RKQ--H 333
QY 292 SCHIPCGKYGKASHLKAHLRWHHTGERFVGNWLCGRFTRSDERHVRTHREKXF 351
Db 334 ICHIPGCEERLYGKASHLKAHLRWHHTGERFLC--LTCGRFRSDELRHGRTHWYRPY 391
QY 352 TCLCSKRFTRSDHLSKHORTH-----GEPGPGPPPSGP 385
Db 392 ACPICKSKFSDHLSKHKHTFKDKKKKVLAAEAKQAAAAIKLEKKEKSKGKPLTPP 451
QY 386 RELGGRSVGEEANOPP-----RSSTSPAPP 412
Db 452 VEFKQEQDPTPLVNYAPYANLYQHSTASAGSSVNPPPP 489

RESULT 9
ABG62110
ID ABG62110 standard; Protein: 612 AA.
XX
AC ABG62110;
XX
19-AUG-2002 (first entry)
XX
Human prostate specific polypeptide #43.
XX
Prostate specific polypeptide; metastasis; prostate cancer; cancer;
KW non-cancerous prostate disease; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200238810-A2.
XX
PD 16-MAY-2002.
XX
06-NOV-2001; 2001WO-US47001.
PR
XX
06-NOV-2000; 2000US-246109P.
PA (DIAD-) DIADEXUS INC.
XX
PI Sun Y, Recipon H, Chen S, Liu C;
```


CC identification of individuals and in forensic analysis, and the
 CC polypeptides may be used as food additives or to prepare antibodies
 CC useful in disease diagnosis, drug targeting and phenotyping. The present
 CC sequence represents a human ovarian antigen of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 135 AA;
 CC Query Match 16.8%; Score 401.5; DB 23; Length 135;
 CC Best Local Similarity 55.7%; Pred. No. 5.8e-22;
 CC Matches 73; Conservative 13; Mismatches 38; Indels 7; Gaps 2;

QY 246 PRGAGTGGGGAGSGAGST---CDPCNQELERLGAAGLKKKPHSCHIPGCGKYV 302
 DB 3 PDPAANGTSPGRDXARGAAPHGCTCPNCKD---GEKRSGEQGGKKHVVCHIPCCKGTF 58
 QY 303 GRASHLKAHLRWHTGPRPVCNWLFCGKRFTSRDELERHVRTHTRKKFTCLLCSKRFTR 362
 DB 59 RKTSLRAHVRUHTGERPVCNWFVCGCKRFTSRDELERHVRTHTRKKFTCLLCSKRFTR 362
 DB 363 SDHLSKHQTH 373
 DB 119 SDHLTRHYKTH 129

RESULT 11
 AAR95975
 ID AAR95975 standard; peptide: 168 AA.
 AC AAR95975;
 DT 19-FEB-1997 (first entry)
 XX Nucleic acid recognition unit #11.
 XX Target binding assembly; nucleic acid recognition unit; NF-kappa-B; SPI;
 KW TATA; human papillomavirus; HPV; HPV E2; human immunodeficiency virus;
 KW HIV LTR; Tat binding unit; probe; assembly sequence; asymmetry sequence;
 KW nuclear localisation signal sequence; human cell; HPV LTR; HIV; bacteria;
 KW virus.
 XX Synthetic.
 XX OS
 XX WO9617956-A2.
 XX PD 13-JUN-1996.
 XX PF 07-DEC-1995; 95WO-US15944.
 XX 09-DEC-1994; 94US-0353476.
 (GENE-) GENE POOL INC.
 Weininger AM, Weininger S;
 WPI; 1996-287199/29.

XX Probe nucleic acids, target binding assemblies, etc - for detection
 PT and localisation of specific nucleic acid sequences, esp. HIV and
 HPV
 XX Claim 14; Page 97-98; 172pp; English.
 XX AAR95965-R95993 represent the nucleic acid recognition units (NAR) of
 CC target binding assemblies (TBA) of the invention. These NARs are
 CC selected from NF-kappa-B, SPI, TATA, human papillomavirus (HPV) E2, HPV
 CC LTR, human immunodeficiency virus (HIV) LTR and Tat binding units. This
 CC sequence represents a SPI sequence NAR. The TBA is recognised by the
 CC target binding region (TBR) of a probe of the invention. The probe of
 CC the invention contains a TBR, a booster binding region (BBR), and an
 CC optional support or attachment (OSA). The TBA contains at least one

CC NAR, and optionally a linker sequence, an assembly sequence, an
 CC asymmetry sequence, a nuclear localisation signal sequence, and an OSA.
 CC The assembly sequence and asymmetry sequences are responsible for the
 CC folding and association of the NARs. The linker sequence is an
 CC oligopeptide, which does not interfere with NAR function, but provides
 CC stability and control over the spacing of the NAR from the rest of the
 CC TBA. The OSA is an attached support or indicator, or other means of
 CC localisation of the probe. The probe can be used in a method for
 CC detecting or localising a specific target nucleic acid sequence (TNA).
 CC The method is highly sensitive, and has a high degree of specificity.
 CC The method can be used for detecting specific nucleic acid sequences,
 CC including those found in human cells, in HIV, HPV, and other nucleic acid
 CC containing systems, including bacteria and viruses.

XX Sequence 168 AA;
 CC Query Match 16.8%; Score 401; DB 17; Length 168;
 CC Best Local Similarity 65.2%; Pred. No. 8.1e-22;
 CC Matches 75; Conservative 6; Mismatches 24; Indels 10; Gaps 2;
 QY 287 KKPIHSHIPGCGKYGVKASHLKAHLRWHTGPRPVCNWLFCGKRFTSRDELERHVRTH 346
 DB 5 KKKQHICHIOGCGKYGVKTSHLRAHLRWHTGPRPVCNWLFCGKRFTSRDELERHVRTH 64
 QY 347 REKFTCLLCSKRFTRSDHLSKHQTH--GEPGPG-----PPSPGPKELGEG 391
 DB 65 GEKFPACPECPKRFMRSDHLSKHQTH--GEPGPG-----PPSPGPKELGEG 391

RESULT 12
 AAR96015
 ID AAR96015 standard; peptide: 241 AA.
 AC AAR96015;
 DT 19-FEB-1997 (first entry)
 XX Target binding assembly #8.
 XX Target binding assembly; nucleic acid recognition unit; NF-kappa-B; SPI;
 KW TATA; human papillomavirus; HPV; HPV E2; human immunodeficiency virus;
 KW HIV LTR; Tat binding unit; probe; assembly sequence; asymmetry sequence;
 KW nuclear localisation signal sequence; human cell; HPV LTR; HIV; bacteria;
 KW virus.
 XX Synthetic.
 XX OS
 XX WO9617956-A2.
 XX PD 13-JUN-1996.
 XX PF 07-DEC-1995; 95WO-US15944.
 XX 09-DEC-1994; 94US-0353476.
 (GENE-) GENE POOL INC.
 Weininger AM, Weininger S;
 WPI; 1996-287199/29.

XX Probe nucleic acids, target binding assemblies, etc - for detection
 PT and localisation of specific nucleic acid sequences, esp. HIV and
 HPV
 XX Claim 25; Page 127-128; 172pp; English.
 XX AAR96008-R96015 represent target binding assemblies (TBA) of the
 CC invention. The TBA is recognised by the target binding region (TBR) of a
 CC probe of the invention. The probe of the invention contains a TBR, a
 CC booster binding region (BBR), and an optional support or attachment
 CC (OSA). The TBA contains at least one nucleic acid recognition unit
 CC (NAR), and optionally a linker sequence, an assembly sequence (see

CC AAR95994-R95998), an asymmetry sequence (see AAR95999-R96006), a nuclear
 CC localisation signal sequence (see AAR96007), and an OSA. The assembly
 CC sequence and asymmetry sequences are responsible for the folding and
 CC association of the NARs. The NARs (see AAR95965-R95993) are selected from
 CC NF-kappa-B, SPI, TATA, human papillomavirus (HPV) E2, HPV LTR, human
 CC immunodeficiency virus (HIV) LTR and Tat binding units. The linker
 CC sequence is an oligopeptide, which does not interfere with NAR function,
 CC but provides stability and control over the spacing of the NAR from the
 CC rest of the TBA. The OSA is an attached support or indicator, or other
 CC means of localisation of the probe. The probe can be used in a method
 CC for detecting or localising a specific target nucleic acid sequence
 CC (TNA). The method is highly sensitive, and has a high degree of
 CC specificity. The method can be used for detecting specific nucleic acid
 CC sequences, including those found in human cells, in HIV, HPV, and other
 CC nucleic acid containing systems, including bacteria and viruses.

XX Sequence 241 AA;

Query Match 16.8%; Score 401; DB 17; Length 241;
 Best Local Similarity 65.2%; Pred. No. 1.2e-21;
 Matches 75; Conservative 6; Mismatches 24; Indels 10; Gaps 2;

QY 287 KPIHSCIPGCGKVGKASHLKAHLRHTGTERPFCVNCWLFCCGRFTRSDLELRHVRHT 346

DB 78 KKKQHICHIQCGKVGKTSHLRAHLRHTGTERPFCVNCWLFCCGRFTRSDLELRHVRHT 137

QY 347 REKFTCLLSKTRTSDHLSKHORTH--GEPGPG-----PPSPGKELGEG 391

DB 138 GERKFAPECPKPRMRSDDLHLSKHKTHQNKKGPGVALSVGTLPLDSCAGSEGG 192

RESULT 13

AAR96011
 ID AAR96011 standard; peptide: 273 AA.

XX AAR96011;

XX 19-FEB-1997 (first entry)

XX Target binding assembly #4.

XX Target binding assembly; nucleic acid recognition unit; NF-kappa-B; SPI;
 KW TATA; human papillomavirus; HPV; HPV E2; human immunodeficiency virus;
 KW HIV LTR; Tat binding unit; probe; assembly sequence; asymmetry sequence;
 KW nuclear localisation signal sequence; human cell; HPV LTR; HIV; bacteria;
 KW virus.

OS Synthetic.

XX WO9617956-A2.

XX 13-JUN-1996.

XX 07-DEC-1995; 95WO-US15944.

XX 09-DEC-1994; 94US-0353476.

XX (GENE-) GENE POOL INC.

XX Weininger AM, Weininger S;

XX WPI; 1996-287199/29.

XX Probe nucleic acids, target binding assemblies, etc - for detection
 PT and localisation of specific nucleic acid sequences, esp. HIV and
 PT HPV

XX Claim 25; Page 121-122; 172pp; English.

XX AAR96008-R96015 represent target binding assemblies (TBA) of the
 CC invention. The TBA is recognised by the target binding region (TBR) of a
 CC probe of the invention. The probe of the invention contains a TBR, a
 CC booster binding region (BBR), and an optional support or attachment

CC (OSA). The TBA contains at least one nucleic acid recognition unit
 CC (NAR), and optionally a linker sequence, an assembly sequence (see
 CC AAR95994-R95998), an asymmetry sequence (see AAR95999-R96006), a nuclear
 CC localisation signal sequence (see AAR96007), and an OSA. The assembly
 CC sequence and asymmetry sequences are responsible for the folding and
 CC association of the NARs. The NARs (see AAR95965-R95993) are selected from
 CC NF-kappa-B, SPI, TATA, human papillomavirus (HPV) E2, HPV LTR, human
 CC immunodeficiency virus (HIV) LTR and Tat binding units. The linker
 CC sequence is an oligopeptide, which does not interfere with NAR function,
 CC but provides stability and control over the spacing of the NAR from the
 CC rest of the TBA. The OSA is an attached support or indicator, or other
 CC means of localisation of the probe. The probe can be used in a method
 CC for detecting or localising a specific target nucleic acid sequence
 CC (TNA). The method is highly sensitive, and has a high degree of
 CC specificity. The method can be used for detecting specific nucleic acid
 CC sequences, including those found in human cells, in HIV, HPV, and other
 CC nucleic acid containing systems, including bacteria and viruses.

XX Sequence 273 AA;

Query Match 16.8%; Score 401; DB 17; Length 273;
 Best Local Similarity 65.2%; Pred. No. 1.4e-21;
 Matches 75; Conservative 6; Mismatches 24; Indels 10; Gaps 2;

QY 287 KPIHSCIPGCGKVGKASHLKAHLRHTGTERPFCVNCWLFCCGRFTRSDLELRHVRHT 346

DB 110 KKKQHICHIQCGKVGKTSHLRAHLRHTGTERPFCVNCWLFCCGRFTRSDLELRHVRHT 169

QY 347 REKFTCLLSKTRTSDHLSKHORTH--GEPGPG-----PPSPGKELGEG 391

DB 170 GERKFAPECPKPRMRSDDLHLSKHKTHQNKKGPGVALSVGTLPLDSCAGSEGG 224

RESULT 14

AAY95986
 ID AAY95986 standard; Protein; 288 AA.

XX AAY95986;

XX 05-DEC-2000 (first entry)

XX Human zinc finger transcription factor RFLAT-1.

XX RFLAT-1; RANTES factor of late activated T-lymphocyte;
 KW transcription factor; zinc finger; human; inflammation;
 KW antiinflammatory; autoimmune disease; AIDS; neoplasia; therapy;
 KW diagnosis.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Domain 1..146

XX /note= "proline-rich transactivation region"

XX Region 147..168

XX /note= "basic region"

XX Domain 169..249

XX /note= "DNA-binding domain, comprises 3 zinc
 fingers"

XX Misc-difference 280

XX /note= "encoded by CCC"

XX WO200052030-A1.

XX 08-SEP-2000.

XX 27-JAN-2000; 2000WO-US02300.

XX 27-JAN-1999; 99US-0117576.

XX (STRD) UNIV LELAND STANFORD JUNIOR.

XX Song AM, Chen Y, Krensky AM;

XX

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model
Run on: February 19, 2003, 13:28:06 ; Search time 47 Seconds
(without alignments)
875.437 Million cell updates/sec

Title: US-09-734-329-2
Perfect score: 2384
Sequence: 1 MASSLLEERAHYGVSSPLAML.....PAPPEKAHGGSPESQNLLEI 428

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

ched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	471	19.8	696	2 A29635	transcription fact
2	466	19.5	788	2 JS0747	regulatory protein
3	463.5	19.4	780	2 A48143	HF-1 regulatory el
4	459.5	19.3	784	2 S26638	SPR-1 protein - hu
5	449.5	18.9	697	2 B44489	GT box-binding pro
6	409	17.2	644	2 S39356	transcription fact
7	402	16.9	495	2 A44489	GT box-binding pro
8	361.5	15.2	382	2 T45072	erythroid Kruppel-
9	356	14.9	233	2 T26781	hypothetical prote
10	347.5	14.6	166	2 T25118	hypothetical prote
11	343.5	14.4	376	2 A48060	erythroid Kruppel-
12	337	14.1	469	2 A57531	EGR alpha transcri
13	335.5	14.1	244	2 I59602	GC box binding pro
14	335.3	14.1	244	2 S25288	BTEB protein - rat
15	332.5	13.9	244	2 JS0748	basic transcriptio
16	320.5	13.4	480	2 S52306	zinc finger protei
17	314.5	13.2	344	2 JC6100	CACCC-box binding
18	314	13.2	309	2 T25800	C2H2-type zinc fin
19	312.5	13.1	419	2 A39892	Wilms' tumor prote
20	307.5	12.9	219	2 S35643	BTEB2 protein - hu
21	306.5	12.9	448	2 S33926	Wilms' tumor prote
22	306	12.8	283	2 JE0235	HIV-promoter GC-ri
23	304.5	12.8	470	2 A30136	developmental cont
24	302	12.7	449	2 A30800	Wilms tumor suscep
25	293	12.3	497	2 T33834	hypothetical prote
26	290.5	12.2	420	2 JC4716	zinc finger DNA-bi
27	290.5	12.2	486	2 A41537	DNA-binding protei
28	289.5	12.1	478	2 A45285	transcription fact
29	288.5	12.1	456	2 A40492	early growth respo

30	288	12.1	410	2 JC5046	Wilms' tumor suppl
31	280	11.7	533	2 JS0304	developmental cont
32	276	11.6	354	2 T22588	hypothetical prote
33	275.5	11.6	445	2 S00256	Krox-20 protein -
34	274.5	11.5	508	2 A32225	nerve growth facto
35	273	11.5	421	2 A56550	Krox-20 - African
36	270.5	11.3	543	2 A41211	early growth respo
37	270.5	11.3	604	2 S05447	finger protein gla
38	268.5	11.3	298	2 T22250	hypothetical prote
39	266.5	11.2	856	2 A39503	B-lymphocyte-induc
40	260	10.9	789	2 A39564	transcription repr
41	259.5	10.9	1173	2 I50620	prockr2 - chicken
42	259	10.9	511	2 I50114	early growth respo
43	257.5	10.8	133	2 I51960	wrl zinc-finger ho
44	257	10.8	485	2 A40751	finger protein MZF
45	256	10.7	447	2 I56511	zic protein - mous

ALIGNMENTS

RESULT 1

A29635
transcription factor Spl - human (fragment)
N:Alternate names: finger protein ZNF76
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 05-Nov-1999
C:Accession: A29635; G44256
R:Kadonaga, J.T.; Garner, K.R.; Mastarz, F.R.; Tjian, R.
Cell 51, 1079-1090, 1987
A:Title: Isolation of cDNA encoding transcription factor Sp1 and functional analysis
A:Reference number: A29635; M01D:88080486; PMID:3319186
A:Accession: A29635
A:Molecule type: mRNA
A:Residues: 1-696 <KAD>
A:Cross-references: GB:J03133; NID:q339517; PID:AAA61154.1; PID:q339518
R:Ragoussis, J.; Senger, G.; Mockridge, I.; Sauseau, P.; Ruddy, S.; Dudley, K.; Shee
Genomics 14, 673-679, 1992
A:Title: A testis-expressed zn finger gene (ZNF76) in human 6p21.3 centromeric to th
A:Reference number: A44256; M01D:93052398; PMID:1427894
A:Accession: G44256
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 569-598 <RAG>
A:Experimental source: T-cell line CEM
A:Note: Sequence extracted from NCBI backbone (NCBIP:125980)
C:Genetics:
A:Gene: GDB:SP1
A:Cross-references: GDB:127453; OMIM:189906
A:Map position: 19q13.1-19q13.3
C:Keywords: DNA binding; transcription regulation; zinc finger

Query Match 19.8%; Score 471; DB 2; Length 696;
Best Local Similarity 45.0%; Pred. No. 3.2e-21;
Matches 107; Conservative 15; Mismatches 78; Indels 38; Gaps 6;

QY	177	GT-LSTGTAQPLNPQLPT-----YPSDFAPLNPAPYPAPHLQPGQHVLPQ	223
Db	425	GTVTVAQAQLSSMPGLQTLNLSALGTSGIQVHTQGLPLAIAANAPGDHGAQLGLHAGGD	484
QY	224	DVYKPAVNGSGOLESGAAKPPRGAGTGGSGGVAGSGAGSTCDPCNCQELERLGRAAA	283
Db	485	GIHDDTAGGEEG--ENSFDAQP-----QAGRRTRREATCTCYCKDSEGGSGDP	531
QY	284	GLRKKPHTSHICPGCKVYGVKASHLKAHLRWHGTGERFVCNWLFCGKRFRTRSDLEHRVR	343
Db	532	G--KKQGHICHQCGKVGKTSHLRAHLRWHGTGERFVCMCTWSYCGKRFTSDLRHQR	589
QY	344	THFREKFTCLLSKRTSRDHLKSHQTH--GEPGPG-----PPSPGPKELGEG	391
Db	590	THTGEKKFACPECPKPRMRSRDLHSKHITKHNKKGPGVALSVGTLPLDSCAGSEGS	647

Nucleic Acids Res. 20, 5519-5525; 1992

A:Title: Cloning by recognition site screening of two novel GT box binding proteins: a f
A:Reference number: S26638; MUID:93087156; PMID:1454515

A:Accession: S26639

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-697 <HAG>

A:Cross-references: EMBL:X69560; NID:938417; PIDN:CAA48562.1; PID:g38418

Query Match

Best Local Similarity 18.9%; Score 449.5; DB 2; Length 697;

Matches 79; Conservative 8; Mismatches 25; Indels 3; Gaps 1;

QY 259 GSGAGRSTCDPCNCELELRLGAAAGLKKPKPIHSCGKVKYKASHLKAHLRWHTGE 318

DB 508 GKRLRVACTCPNCKE---GGGRGTNKGKKQHICHIFCGKVKYKTSHLRAHLRWSGE 564

319 RPFVNCNWLFCGKRFTRSDLELHVHTHREKFTCLLCSKRFTRSDHLSKHORTH 373

DB 565 RPFVNCNMYCGRFRTRSDLELQHRTHYGEKKFVCEPKRPMRSDHLAKHIKH 619

RESULT 6

S39356

transcription factor btd - fruit fly (*Drosophila* sp.)

C:Species: *Drosophila* sp.

C:Date: 18-Feb-1994 #sequence_revision 01-Dec-1995 #text_change 07-May-1999

A:Accession: S39356

R:Wimmer, E.A.; Jaecckle, H.; Pfeifle, C.; Cohen, S.M.

Nature 366, 690-694, 1993

A:Title: A *Drosophila* homologue of human Spl is a head-specific segmentation gene.

A:Reference number: S39356; MUID:94081952; PMID:8259212

A:Accession: S39356

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-644 <WIN>

A:Cross-references: EMBL:Z29361; NID:9441283; PID:g441284

C:Genetics:

A:Gene: FlyBase:btd

A:Cross-references: FlyBase:FBgn0000233

A:Introns: 245/2

Query Match

Best Local Similarity 17.2%; Score 409; DB 2; Length 644;

Matches 132; Conservative 34; Mismatches 122; Indels 170; Gaps 18;

65 FSTNGLLSP-----AGSPAPASGYANDY--PPFPHSPGP-----TGAQDPCGLLPKQHS 114

DB 92 FLTAAALLSAPSLSGSSSGSSGSSPLYKPKMLPLPYQASSTGTASPNSSIQSAPS 151

QY 115 SSDCLPSVYTSUDMTHPYGSKYKAGIHAGISFGPGNTPTPMDMHPGNNLWGGGQGDG 174

DB 152 SASVSPIPS-----PAQSFASISASP-STPTT----- 179

QY 175 LOGTLSTGAQPLNP---QLPTYSDFAPLNPAPYPAPHLLQPCQHVLPQDVYKPAV 231

DB 180 ----TLA-----PPTTAAGALAGSTSSPSSSASAA-----AAAAAA 216

QY 232 GNSGLEGSGAAKPPRGAGTGGG-----GVAGSGAGRST----- 266

DB 217 AAADLGAANAASAYGNWNTAYSGLGPARSOPFYAQYASDIYGNVAGMSSAAWFSHER 276

QY 267 -----CDCPNC-QELERL-----GAAAGLRKKPIH 291

DB 277 LYQWSSQSYPGNFDDIAFOTOLQRRSVRCCTPCNTNEMSGLPPIVGPDERG-RKQ--H 333

QY 292 SCHIPGCKYKASHLKAHLRWHTGERPFCVNCNWLFCGKRFTRSDLELHVHTHREKFF 351

DB 334 ICHIPGCELYGKASHLKAHLRWHTGERPFLC--LTCGKRFTRSDLELQHRGHTHTNYRYP 391

QY 352 TCLLCSKRFTRSDHLSKHORTH-----GEPGPGPPSPG 385

DB 392 ACPICSKFRSDHLSKHKKHFKDKSKVLAABEQAQAAAAIKLEKKEKSGKPLTPP 451

QY 386 KELGEGSRVGEAEANOPP-----RSSTSPAPP 412

DB 452 VEFKQEQDPTTPLVNYAPYANLYQHTSAGSSVNPPPP 489

RESULT 7

A44489

GT box-binding protein Sp2 - human

N:Alternate names: transcription factor Sp2

C:Species: *Homo sapiens* (man)

C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999

C:Accession: A44489

R:Kingsley, C.; Winoto, A.

Mol. Cell. Biol. 12, 4251-4261, 1992

A:Title: Cloning of GT box-binding proteins: a novel Spl multigene family regulation

A:Reference number: A44489; MUID:93024366; PMID:1341900

A:Accession: A44489

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-495 <KIN>

A:Cross-references: GB:M97190; NID:g338300; PIDN:AAA36629.1; PID:g338301

A:Experimental source: Molt13

A:Note: sequence extracted from NCBI backbone (NCBIP:114078)

C:Keywords: DNA binding; transcription regulation

Query Match

Best Local Similarity 16.9%; Score 402; DB 2; Length 495;

Matches 101; Conservative 20; Mismatches 71; Indels 90; Gaps 10;

QY 178 TLSTGAPQPLNPQL-----PTY-----PS-----DFAPLNPAPYP----- 208

DB 212 TLPTVPQKPSQNFQIAAEPTPTQVYIRTPSGEVOTLVQDPPATAAATNTTSSPAS 271

QY 209 -APHLLOPQPOH---VLQDVYKPAVG-----NSGOLESGAAK----- 244

DB 272 RAPLUSGTSKKHSAAILRKRPKPIAPAGSIISLNAQAALAAQAAMOTININGVOVG 331

QY 245 PPRGAGTGG-----SGG-----YAGSGAGRSTCDCPN 271

DB 332 PVTINTGQQQLTVQNVSGNNLTISGLSPTQIQLOMEQALAGETQPGKRRRMACTCPN 391

QY 272 COELERLGAAGAAGLKKPKPIHSCGKVKYKASHLKAHLRWHTGERPFCVNCNWLFCGKR 331

DB 332 CKD-----GKRSGEQGGKKKHVCHIPDCGKTRFKTSLRAHVLHTGERPFCVNCNWLFCGKR 447

QY 332 FTRSDLELHVHTHREKFTCLLCSKRFTRSDHLSKHORTH 373

DB 448 FTRSDLELQHRHARTHTGDKRFECAQCOKRPMRSDHLTKHYKTH 489

RESULT 8

T45072

erythroid Kruppel-like factor homolog [imported] - human

C:Species: *Homo sapiens* (man)

C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000

C:Accession: T45072

R:Lamerdin, J.; McCreedy, P.; Stilwagen, S.; Ramirez, M.; Carrano, A.

submitted to the EMBL Data Library, November 1996

A:Description: Characterization by genomic sequence analysis of a gene-rich 111 kb

A:Reference number: 222906

A:Accession: T45072

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-362 <LAM>

A:Cross-references: EMBL:AD000092; PIDN:AAB51173.1

A:Experimental source: cell line 5HL2-B; fibroblast

C:Genetics:

A:Map position: 19p13.2

A:Introns: 29/3; 305/1

A:Note: EKL

Query Match

15.2%; Score 361.5; DB 2; Length 362;

```

T25118
hypothetical protein T22C8.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T25118
R:Thomas, K.
submitted to the EMBL Data Library, April 1995
A:Reference number: Z19983
A:Accession: T25118
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-166 <MIL>
A:Cross-references: EMBL:Z49071; PIDN:CAA88877.1; GSPD:GN00020; CESP:T22C8.5
A:Experimental source: clone T22c8
C:Genetics:
A:Gene: CESP.T22C8.5
A:Map position: 2
A:Introns: 66/2; 99/2; 136/2

Query Match          14.6%; Score 347.5; DB 2; Length 166;
Best Local Similarity 48.5%; Pred. No. 2.2e-14;
Matches 65; Conservative 16; Mismatches 38; Indels 15; Gaps

QY 245 PPRGAGTGSGGGYAGSAGRGST-----CDCPNCQELERLGAAAGLR-KKPIHSCHI 295
      | | | | | | | | | | | | | | | | | | | | | | : | | | | |
Db 28 PDSPASTASSSSSIGANELLTKRRKCERCCTCPNKAIKH-----GDGRGSOHTHLCV 81

QY 296 PGCGKYVGKASHLKAHLRWHTGERPFVCWNWFCGKRFTSRDELERHVTRTHREKKFTCLL 355
      | | | | | | | | | | | | | | | | | | | | | | : | | | | |
Db 82 PGCGKYTKTSHLRAHLRKHTGDRPFVDFDCGKRFRSDOLIRHKRTHTEYRFACKF 141

QY 356 CSKRFTSRDHLSKH 369
      | | | | | | | | | | | | | | | | | | | | | | : | | | | |
Db 142 CIRQFSRDHLQQH 155

RESULT 11
A48060
erythroid Kruppel-like factor EKLF - mouse
C:Species: Mus musculus (house mouse)
C:Date: 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C:Accession: A48060
R:Miller, I.J.; Bleker, J.J.
Mol. Cell. Biol. 13, 2776-2786, 1993
A>Title: A novel, erythroid cell-specific murine transcription factor that blin
A:Reference number: A48060; MUID:93233640; PMID:7682653
A:Accession: A48060
A>Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-376 <MIL>
A:Cross-references: GB:M97200; NTD:gl93011; PIDN:AAA37546.1; PID:g193012
A>Note: sequence extracted from NCBI backbone (NCBIN:129821, NCBI:P:129822)

Query Match          14.4%; Score 343.5; DB 2; Length 376;
Best Local Similarity 30.6%; Pred. No. 8.4e-14;
Matches 116; Conservative 32; Mismatches 117; Indels 114; Gaps 2

QY 97 PGPTGAODPLLVPKGHSDDCLPSVYT-----SLDMTHPYGSWKAGIHAGISPGPN 150
      | | | | | | | | | | | | | | | | | | | | | | : | | | | |
Db 9 PEVGGGHQPANA-----SAETVLPSISTLTTLTGOFLDTQEDFLKWWRSEETQDLGPGPN 63

QY 151 TPFPWDMDH-----PGGN-----W-----LGGGGQGGLQGTLL--STGPA 184
      | | | | | | | | | | | | | | | | | | | | | | : | | | | |
Db 64 PTGP--SLHVSLSKEDPSGEDDERDYTCAWDPDLTFNFGSESPTGRTCALAPSVGPV 121

QY 185 ---QPLEN-----POLPTYP-----SDEFA-----PLNPAPYPAPHLLQPG---POHVL 221
      | | | | | | | | | | | | | | | | | | | | | | : | | | | |
Db 122 AQEPPPSLCAAYAGGGLVTGPLGSEHTSWAUPTRPPAPEPVFVAP-ALAPGLAKP--- 177

QY 222 PQDIYKPAYGN-SGQLEGSGAAKPPR-GAGTGGSGGY----- 257
      | | | | | | | | | | | | | | | | | | | | | | : | | | | |
Db 178 AOPSYSDSRAGSVGGFFPRAGLAVPAAPCAPYGLLSGYPALYPAPQYQHQLFRCLAAP 237

```



```
Db 104 GSPSPHSPEERQDSG-----APSPLSLLHSGVASKGHASEKRRHKCPYSG 149
QY 298 CGKVYKASHLKAHLRWHTGERPFVCNWLFCGKRETRSDLELHRYHTTREKKFTCLCS 357
Db 150 CGKVYKSSHLKAHYRVHTGERPFPCTWPDCLKFSRDELTRHYRTHTGEXQRCPLCE 209
QY 358 KRFRSDHLSKHQTHGEPG 378
Db 210 KRFRSDHLTKHARRHDFHP 230
```

Search completed: February 19, 2003, 13:32:24
Job time : 49 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 19, 2003, 13:21:36 ; Search time 31 Seconds
(without alignments)
572.641 Million cell updates/sec

Title: US-09-734-329-2

Perfect score: 2384

Sequence: 1 MASSLLEEAHYGSSPLAML.....PAPPEKAHGSPQSNLLEI 428

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

..rched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	473	19.8	152	1 SP6_MOUSE	Q9esx2 mus musculus
2	471	19.8	785	1 SP1_HUMAN	P08047 homo sapien
3	466	19.5	788	1 SP1_RAT	Q01714 rattus norv
4	459.5	19.3	784	1 SP4_HUMAN	Q02446 homo sapien
5	449.5	18.9	711	1 SP3_HUMAN	Q02447 homo sapien
6	411	17.2	289	1 KLFD_MOUSE	Q9fjz6 mus musculus
7	409	17.2	644	1 BTD_DROME	Q24266 drosophila
8	406	17.0	606	1 SP2_HUMAN	Q02086 homo sapien
9	398.5	16.7	252	1 BTE4_HUMAN	Q9bxi1 homo sapien
10	395.5	16.6	288	1 KLFD_HUMAN	Q9y2y9 h krueppel-
11	389.5	16.3	416	1 KLFD_HUMAN	Q9uih9 homo sapien
12	384.5	16.1	354	1 KLFD_MOUSE	P58334 mus musculus
13	379	15.9	251	1 BTE4_MOUSE	Q9et58 rattus norv
14	374	15.7	351	1 KLFD_RAT	Q9epx2 mus musculus
15	364	15.3	415	1 KLFE_MOUSE	Q13351 homo sapien
16	361.5	15.2	362	1 KLFD_HUMAN	Q9y5w3 homo sapien
17	359.5	15.1	355	1 KLFD_HUMAN	Q14901 homo sapien
18	355	14.9	512	1 KLFB_MOUSE	Q9epf4 mus musculus
19	354	14.8	511	1 KLFB_MOUSE	Q04374 homo sapien
20	353.5	14.8	470	1 KLFD_HUMAN	P46059 mus musculus
21	343.5	14.4	358	1 KLFL_MOUSE	Q08876 rattus norv
22	343.5	14.4	480	1 KLFA_RAT	Q06793 mus musculus
23	339.5	14.2	474	1 KLFA_MOUSE	Q89091 mus musculus
24	337.5	14.2	479	1 KLFA_MOUSE	Q13118 homo sapien
25	337	14.1	480	1 KLFA_HUMAN	Q13886 homo sapien
26	335.5	14.1	244	1 BTEL_HUMAN	Q01713 rattus norv
27	332.5	13.9	244	1 BTEL_RAT	Q35739 mus musculus
28	331.5	13.9	244	1 BTEL_MOUSE	Q13887 homo sapien
29	323	13.5	457	1 KLFS_HUMAN	Q92027 mus musculus
30	315.5	13.2	446	1 KLFS_MOUSE	Q60980 mus musculus
31	314.5	13.2	344	1 KLFD_MOUSE	Q9udv7 homo sapien
32	311	13.0	671	1 Z282_HUMAN	P22561 mus musculus
33	309.5	13.0	449	1 WTL_MOUSE	

34	306.5	12.9	448	1 WTL_RAT	P49952 rattus norv
35	306	12.8	283	1 KLFD_HUMAN	Q99612 homo sapien
36	306	12.8	449	1 WTL_PIG	O62651 sus scrofa
37	305	12.8	283	1 KLFD_RAT	Q35819 rattus norv
38	305	12.8	359	1 KLFD_HUMAN	O95600 homo sapien
39	304.5	12.8	470	1 EGR2_MOUSE	P08152 mus musculus
40	302	12.7	283	1 KLFD_MOUSE	O08584 mus musculus
41	302	12.7	449	1 WTL_HUMAN	P19544 homo sapien
42	299.5	12.6	301	1 KLFD_MOUSE	Q99jbo mus musculus
43	293	12.3	402	1 KLFD_MOUSE	Q35738 mus musculus
44	293	12.3	734	1 ZN42_HUMAN	P28698 homo sapien
45	290.5	12.2	470	1 EGR2_RAT	P51774 rattus norv

ALIGNMENTS

RESULT 1

SP6_MOUSE

ID SP6_MOUSE STANDARD; PRT; 152 AA.

AC Q9ESX2;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Transcription factor Sp6 (Krueppel-like factor 14) (Fragment).

GN SP6 OR KLF14.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20541714; PubMed=11087666;

RA Scohy S., Gabant P., Van Reeth T., Hertveldt V., Dreze P.-L.,

RA Van Vooren P., Riviere M., Szpirer J., Szpirer C.;

RT Identification of KLF13 and KLF14 (SP6), novel members of the SP/XKLF

RT Transcription factor family.;

RL Genomics 70:93-101(2000).

CC -1- TISSUE SPECIFICITY: Ubiquitous.

CC -1- SIMILARITY: BELONGS TO THE SP1 FAMILY OF C2H2-TYPE ZINC-FINGER

CC PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AJ275988; CAC06698.1; .

DR HSP; P08047; ISP2.

DR MGD; MGI:1932575; Sp6.

DR InterPro; IPR000822; Znf_C2H2.

DR Pfam; PF00096; zf-C2H2; 3.

DR ProDom; PD000003; Znf_C2H2; 1.

DR SMART; SM00355; Znf_C2H2; 3.

DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.

DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.

KW Transcription regulation; DNA-binding; Nuclear protein; Repeat;

Zinc-finger; Metal-binding.

FT NON_TER 1 1

FT DOMAIN 30 112 ZINC FINGERS.

FT 2N_FING 30 54 C2H2-TYPE.

FT 2N_FING 60 84 C2H2-TYPE.

FT 2N_FING 90 112 C2H2-TYPE.

SQ SEQUENCE 152 AA; 16456 MW; AF629C4845599938 CRC64;

Query Match 19.8%; Score 473; DB 1; Length 152;

Best Local Similarity 58.68; Pred. No. 5.2e-20;

Matches 85; Conservative 17; Mismatches 29; Indels 14; Gaps 3;

QY 264 RSTCDPCNCELERLGAAA--GLRKKPIHSCHPGKGVYKASHLKAHLRWHHTGERPF 321
 Db 1 QTVCRNCLDAERLAPGDPGGKKKHLNCHIPGCGKAYAKTSHLKAHLRWHSGDRPF 60
 QY 322 VCNLFQCGKFRTRDELEHVRHTRKKTCLCSKFRTRSDHLSKHQTHGEPGPGPP 381
 Db 61 VCNLFQCGKFRTRDELEHVRHTRKKTCLCSKFRTRSDHLSKHQTHGEPGPGPP 112
 QY 382 PSQPKELGGRSVGEERAN---QPP 403
 Db 113 -EKAKEERAAAQGEKAGGVVEPP 136

RESULT 2
 SPL_HUMAN STANDARD; PRT; 785 AA.
 ID SPL_HUMAN AC P08047; Q9NVE7; Q9H3Q5;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Transcription factor Spl.
 GN SPL OR TSFPI.
 Homo sapiens (Human).
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 [1]
 RN SEQUENCE OF 4-785 FROM N.A.
 RP TISSUE=Cervical carcinoma;
 RA Heggart M.H., Ladurner A.G.;
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 [2]
 RN SEQUENCE OF 1-558 FROM N.A.
 RP MEDLINE=20545561; PubMed=10973950;
 RA Takahara T., Kanazu S., Yanagisawa S., Akanuma H.;
 RT "Heterogeneous Spl mRNAs in human HepG2 cells include a product of
 RT homotypic trans-splicing";
 RL J. Biol. Chem. 275:38067-38072(2000).
 [3]
 RN SEQUENCE OF 90-785 FROM N.A., AND SEQUENCE OF 359-375 AND 670-675.
 RP MEDLINE=88080466; PubMed=3319186;
 RA Kadonaga J.T., Carner K.R., Maslarz F.R., Tjian R.;
 RT "Isolation of cDNA encoding transcription factor Spl and functional
 RT analysis of the DNA binding domain.";
 RL Cell 51:1079-1090(1987).
 [4]
 RN O-GLYCOSYLATION.
 RP MEDLINE=89003041; PubMed=3139301;
 RA Jackson S.P., Tjian R.;
 RT "O-glycosylation of eukaryotic transcription factors: implications
 RT for mechanisms of transcriptional regulation.";
 RL Cell 55:125-133(1988).
 [5]
 RN STRUCTURE BY NMR OF 654-684 AND 684-712.
 RP MEDLINE=97218212; PubMed=9065444;
 RA Narayan V.A., Kriwacki R.W., Caradonna J.P.;
 RT "Structures of zinc finger domains from transcription factor Spl.
 RT Insights into sequence-specific protein-DNA recognition.";
 RL J. Biol. Chem. 272:7801-7809(1997).
 [6]
 RN IDENTIFICATION OF SEROTONIN 1A RECEPTOR PROMOTER BINDING SITES.
 RP MEDLINE=96224025; PubMed=8626793;
 RA Parks C.L., Shenk T.;
 RT "The serotonin 1a receptor gene contains a TATA-less promoter that
 RT responds to MAZ and Spl.";
 RL J. Biol. Chem. 271:4417-4430(1996).
 CC -1- FUNCTION: BINDS TO GC BOX PROMOTERS ELEMENTS AND SELECTIVELY
 CC ACTIVATES MRNA SYNTHESIS FROM GENES THAT CONTAIN FUNCTIONAL
 CC RECOGNITION SITES. CAN INTERACT WITH G/C-RICH MOTIFS FROM
 CC SEROTONIN RECEPTOR PROMOTER.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- PTM: O-GLYCOSYLATED; CONTAINS N-ACETYLGLUCOSAMINE SIDE CHAINS.
 CC -1- SIMILARITY: BELONGS TO THE SPL FAMILY OF C2H2-TYPE ZINC-FINGER

PROTEINS.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; AF252284; AAF67726.1; -
 DR EMBL; AB039286; BAB13476.1; -
 DR EMBL; J03133; AAA61154.1; -
 DR FIR; A29635; A29635.
 DR PDB; 1SP1; 21-APR-97.
 DR PDB; 1SP2; 21-APR-97.
 DR TRANSFAC; T00759; -
 DR GlycoSuiteDB; P08047; -
 DR Genew; HGNC:11205; SPL.
 DR MIM; 189906; -
 DR InterPro; IPR000822; Znf_C2H2.
 DR Pfam; PF00096; zf-C2H2; 3.
 DR ProDom; PD000003; Znf_C2H2; 2.
 DR SMART; SM00355; Znf_C2H2; 3.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
 DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 3.
 KW DNA-binding; Nuclear protein; Repeats; Zinc-finger; Metal-binding;
 FT DOMAIN 626 708 ZINC FINGERS.
 FT ZN_FING 626 650 C2H2-TYPE.
 FT ZN_FING 656 680 C2H2-TYPE.
 FT ZN_FING 686 708 C2H2-TYPE.
 FT CONFLICT 366 366 D -> G (IN REF. 3; AA SEQUENCE).
 FT CONFLICT 670 670 S -> F (IN REF. 3; AA SEQUENCE).
 SQ SEQUENCE 785 AA; 80693 MW; 43893DBF6518B9EA CRC64;

Query Match 19.8%; Score 471; DB 1; Length 785;
 Best Local Similarity 45.0%; Pred. No. 3e-19;
 Matches 107; Conservative 15; Mismatches 78; Indels 38; Gaps 6;

QY 177 GTLSTGPAQPLNPQLPT-----YPSDFAPLNPAPYPAPPHLQPGQHVLPQ 223
 Db 514 GTVYVNAQLSSHPGLQTLNLSALGTSIGVHPQGLPLAANAPGHHQLGHHAGGD 573
 QY 224 DVTYKRAVNGSGOLESGAAKPPRGACTGGGGYAGSGACRSTCDPCNCELERLGAAA 283
 Db 574 GIHDDTAGGEG--ENSPDAQP-----QAGRRTRREACTCPYCKDSEGRGSGDP 620
 QY 284 GLRKKPIHSCHPGKGVYKASHLKAHLRWHHTGERPFVNCWLFQCGKFRTRSDLEHVR 343
 Db 621 G--KKQKHICHQCGKGVYKGTSHLRAHLRWHHTGERPFVNCWLFQCGKFRTRSDLEHVR 678
 QY 344 THTRKKFTCLCSKFRTRSDHLSKHQTHGEPGPG-----PPSPGPKELGEG 391
 Db 679 THTRKKFTCLCSKFRTRSDHLSKHQTHGEPGPG-----PPSPGPKELGEG 736

RESULT 3

SPL_RAT STANDARD; PRT; 788 AA.
 ID SPL_RAT AC Q01714;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Transcription factor Spl.
 GN SPL.
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 CC NCBI_TaxID=101116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;

Db 649 CHIEGCKVYGTSTLRAHLRHTGEPFICNMFCGRFTRSDLOQRHRTHTGKRF 708
QY 353 CLLCCKRFRSDHLSKHORTHCPEGPPSPGKELGGRSVGEAEANQPPRSSTSPAPP 412
Db 709 CPECCKRFRSDHLSKHVYKTHQNKGGGTALAVTSGELDS-SVTEVLGSPRIVTVAIS 767
QY 413 ERAHGSPQSNLE 427
Db 768 QDSNPATPNVSTNME 782

RESULT 5
SP3_HUMAN STANDARD; PRT; 711 AA.
AC Q02447;
DT 01-FEB-1995 (Rel. 31, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcription factor Sp3 (SP3-2) (Fragment).
GN SP3.
OS Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A., AND REVISIONS.
RC TISSUE=T-cell;
RA Kingsley C., Winoto A.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE OF 59-711 FROM N.A.
RX MEDLINE=93024366; PubMed=1341900;
RA Kingsley C., Winoto A.;
RT "Cloning of GT box-binding proteins: a novel Sp1 multigene family
regulating T-cell receptor gene expression.";
RL Mol. Cell. Biol. 12:4251-4261(1992).
RN [3]
SEQUENCE OF 15-711 FROM N.A.
RC TISSUE=Uterus;
RX MEDLINE=93087156; PubMed=1454515;
RA Hagen G., Mueller S., Beato M., Suske G.;
RT "Cloning by recognition site screening of two novel GT box binding
proteins: a family of Sp1 related genes.";
RL Nucleic Acids Res. 20:5519-5525(1992).
RN [4]
FUNCTION.
RX MEDLINE=97426517; PubMed=9278495;
RA Ihn H., Trojanowska M.;
RT "Sp3 is a transcriptional activator of the human alpha2(I) collagen
gene.";
CC Nucleic Acids Res. 25:3712-3717(1997).
CC -!- FUNCTION: BINDS TO GT AND GC BOXES PROMOTERS ELEMENTS. PROBABLE
TRANSCRIPTIONAL ACTIVATOR.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED.
CC -!- SIMILARITY: BELONGS TO THE SP1 FAMILY OF C2H2-TYPE ZINC-FINGER
PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M97191; AAA36630.2; -
CC EMBL; X68560; CAA48562.1; -
CC PIR; S26639; S26639.
CC HSSP; P08047; 1SP1.
CC TRANSFAC; T02338; -
CC Genew; HGNC:11208; SP3.

MM; 601804; -
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 3.
DR ProDom; PD000003; Znf_C2H2; 1.
DR SMART; SM00355; Znf_C2H2; 3.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 3.
KW Transcription regulation; Activator; Zinc-finger; Metal-binding;
KW DNA-binding; Nuclear protein; Repeat.
FT NON_TER 1
FT DOMAIN 551 633 ZINC_FINGERS.
FT ZN_FING 551 575 C2H2-TYPE.
FT ZN_FING 581 605 C2H2-TYPE.
FT ZN_FING 611 633 C2H2-TYPE.
FT CONFLICT 94 94 T -> A (IN REF. 3).
FT CONFLICT 669 669 N -> K (IN REF. 3).
SQ SEQUENCE 711 AA; 75102 MW; D1719FAC9B05A217 CRC64;
Query Match 18.9%; Score 449.5; DB 1; Length 711;
Best Local Similarity 68.7%; Pred. No. 4.1e-18;
Matches 79; Conservative 8; Mismatches 25; Indels 3; Gaps 1;
QY 259 GSGAGRSTCDPCNCOELERLGNAAAGLRKPHSCHIPGCGKVKYKASHLKAHLRWHTGE 318
DB 522 GKRLRRVACTCPCKE---GGGRGTLGKKQKHICHIPGCGKVKYKTSHLRAHLRWSGE 578
QY 319 RPFVCNWLFCGKRFTRSDLEHVRHTREKFTCLLCCKRFRSDHLSKHORTH 373
DB 579 RPFVCNWLFCGKRFTRSDLEHVRHTREKFTCLLCCKRFRSDHLSKHORTH 633
RESULT 6
KLFD_MOUSE STANDARD; PRT; 289 AA.
ID KLFD_MOUSE
AC Q9JZ6; O9JH8; O9ESX3;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Krueppel-like factor 13 (transcription factor BTEB3) (Basic
transcription element binding protein 3) (BTE-binding protein 3)
DE (RANTES factor of late activated T lymphocytes-1) (RFLAT-1) (Erythroid
transcription factor FKLF-2).
DE KLF13 OR BTEB3 OR FKLF2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=20115087; PubMed=10642511;
RA Martin K.M., Cooper W.N., Metcalfe J.C., Kemp P.R.;
RT "Mouse BTEB3, a new member of the basic transcription element binding
protein (BTEB) family, activates expression from GC-rich minimal
promoter regions.";
RT Biochem. J. 345:529-533(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=B6D2F2; TISSUE=Yolk;
RX MEDLINE=20287401; PubMed=10828046;
RA Asano H., Li X.S., Stamatoyannopoulos G.;
RT "FKLF-2, a novel Krueppel-like transcriptional factor that activates
globin and other erythroid lineage genes.";
RT Blood 95:3578-3584(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal liver;
RX MEDLINE=20541714; PubMed=11087666;
RA Scohy S., Gabant P., Van Reeth T., Hertveldt V., Dreze P.-L.,
RA Van Vooren P., Riviere M., Szpirer J., Szpirer C.;
RT "Identification of KLF13 and KLF14 (SP6), novel members of the SP/XKLF
transcription factor family.";
RL Genomics 70:93-101(2000).

CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- DEVELOPMENTAL STAGE: FIRST EXPRESSED IN A STRIPE COVERING THE HEAD
CC ANLAGEN OF THE SYNCHITIAL BLASTODERM EMBRYO, PERSISTS THROUGH
CC GASTRULATION AND DECAYS DURING GERM BAND EXTENSION. EXPRESSED
CC LATER IN DEVELOPMENT IN A COMPLEX SPATIALLY RESTRICTED PATTERN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z29361; CA82545.1; .
DR EMBL; AE003448; AF46518.1; .
DR HSSP; P08047; LSP2
DR FlyBase; FBgn000233; btd.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; Znf_C2H2_3.
DR PRINTS; PR00048; ZINCFINGER.
DR PRODOM; PD000003; Znf_C2H2_1.
DR SMART; SM00355; Znf_C2H2_3.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 3.
KW Transcription regulation; Activator; Zinc-finger; Metal-binding;
KW DNA-binding; Nuclear protein; Repeat.
FT DOMAIN 333 413 ZINC FINGERS.
FT ZN_FING 333 357 C2H2-TYPE.
FT ZN_FING 363 385 C2H2-TYPE.
FT ZN_FING 391 413 C2H2-TYPE.
FT DOMAIN 14 89 GLN-RICH.
FT DOMAIN 208 220 POLY-ALA.
FT DOMAIN 431 434 POLY-ALA.
FT DOMAIN 486 492 POLY-PRO.
FT DOMAIN 499 502 POLY-THR.
FT DOMAIN 515 519 POLY-SER.
FT DOMAIN 530 536 POLY-SER.
FT DOMAIN 596 599 POLY-SER.
SQ SEQUENCE 644 AA; 68581 MW; A0DB98C2AF938452 CRC64;

Query Match 17.28; Score 409; DB 1; Length 644;
Best Local Similarity 28.88; Pred. No. 6.2e-16;
Matches 132; Conservative 34; Mismatches 122; Indels 170; Gaps 18;

QY 65 FSSTNGLLSP---AGSPAPASGVANDY---PPPHSPFGP---TGAQDPGLLVPRGHS 114
DB 92 FLSSAALLSAPPSLSGSSGSSGSSPLYGKPKKLELPYQASSTGTASPNSIQSAPS 151
QY 115 SSDCLPSVYTSLDMTHPYGSWKAGIHAGISPGFGCNTPTPWMDHPGNGWLGQGGGQGDG 174
152 SASVSPSIFPS-----PAQSFAISASP-STPTT----- 179
QY 175 LQGLTSTGAPPLNP---QLPTTSPDAPLNPAPYPAPHLLQPGQHVLPQDVKPKAV 231
DB 180 ---TLA-----PPTAAGALAGSPTSSPSSSAASA-----AAAAAA 216
QY 232 NSGQLEGSGAAPPKRGAGTGGSG-----GYAGSGAGST----- 266
DB 217 AAADLGAARAVASAYGNWTAISGLGPAPRSOPFYAQVASYDIGNAVGSSAAWFSQER 276
QY 267 -----CDPCNC-OELERL-----GAAAAGLRKKPIH 291
DB 277 LVQPMSSQSYGFNFDDIAFOTLQRRSVRCCTPCNCTNEMSGLPPIVGPDERG-RKQ--H 333
QY 292 SCHTPGCKVYKASHLKAHLRWHTGTGPPVFCVNMFLFCGKRTSRDELRHVRTHTRKKF 351
DB 334 ICHTPGCRLYGKASHLKAHLRWHTGTGPPVFCVNMFLFCGKRTSRDELRHVRTHTRKKF 391
QY 352 TCLLCSKRTSRDLHLSKHQTH-----GEPGCPGPPSGP 385
DB 392 ACPICSKAFSRDLHLSKHKKTHFKDKKSKVLAARKEQAAAAIKLEKKKSGKPLTTP 451

QY 386 KELGGRSVGEEANQPP-----RSTSPAPP 412
DB 452 VEFKQEQDTPPLVYAPYANLYQHSISAGSSVNPAPP 489
RESULT 8
SP2_HUMAN STANDARD; PRT; 606 AA.
ID SP2_HUMAN
AC Q02086;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcription factor Sp2.
GN SP2 OR KIAA0048.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo;
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Bone marrow;
RX MEDLINE-96051398; Pubmed=7584044;
RA Nomura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S.,
RA Seki N., Kawarabayashi Y., Ishikawa K.-I., Tabata S.;
RT "Prediction of the coding sequences of unidentified human genes. II.
RT The coding sequences of 40 new genes (K1AA0041-K1AA0080) deduced by
RT analysis of cDNA clones from human cell line KG-1."
RL DNA Res. 1:223-229(1994).
RN [2]
RP SEQUENCE OF 112-606 FROM N.A.
RX MEDLINE-93024366; Pubmed=1341900;
RA Kingsley C., Winoto A.;
RT "Cloning of GT box-binding proteins: a novel Sp1 multigene family
RT regulating T-cell receptor gene expression."
RL Mol. Cell. Biol. 12:4251-4261(1992).
CC -!- FUNCTION: BINDS TO GC BOX PROMOTERS ELEMENTS AND SELECTIVELY
CC ACTIVATES MRNA SYNTHESIS FROM GENES THAT CONTAIN FUNCTIONAL
CC RECOGNITION SITES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE SP1 FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D28588; BAA05923.1; .
DR EMBL; M97190; AAA36629.1; .
DR PIR; A44489; A44489.
DR HSSP; P08047; LSP2.
DR TRANSFAC; T02356; .
DR Genew; HGNC:11207; SP2.
DR MIM; 601801; .
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; Znf_C2H2_3.
DR PRINTS; PR00048; ZINCFINGER.
DR PRODOM; PD000003; Znf_C2H2_2.
DR SMART; SM00355; Znf_C2H2_3.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 3.
KW Transcription regulation; Activator; Zinc-finger; Metal-binding;
KW DNA-binding; Nuclear protein; Repeat.
FT DOMAIN 518 600 ZINC FINGERS.
FT ZN_FING 518 542 C2H2-TYPE.
FT ZN_FING 548 572 C2H2-TYPE.
FT ZN_FING 578 600 C2H2-TYPE.
SQ SEQUENCE 606 AA; 64153 MW; A27C6D460D36E186 CRC64;

Query Match 17.08; Score 406; DB 1; Length 606;

Best Local Similarity 25.38; Pred. No. 8.5e-16;
Matches 153; Conservative 37; Mismatches 139; Indels 276; Gaps 24;

```
QY 15 SPLAMLTAACSKFGSGSPLRSTLTGKGKTKKPYADLSAPKTMGDAY--PAPESTNGILL 72
      |||||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 26 SPLALLAATCSKIG-----PPAVEAAVTTPAPQPTPRKL 60

QY 73 SPAGSPAPASGANDY-----PPPHSFGPPT---GAQDPGLLVKPGHSS 115
      |||||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 VPIKPAPLPLSPGKNSFGILSKGNILQIOGSOLSASYPGQLVFAIQNP-TWINKGTRS 119
      |||||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 116 S-----DCLPSVYISLDMT-----HPYGSWYK----- 137
      |||||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 120 NANOIQAVPQIQASNSGTIOQPNLNLQIIPGTNQAIITPSSHKPVPIKPAPIQK 179
      |||||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 138 -----AGIHAG-----ISPGPGNT-----PTPMD----- 157
      |||||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 180 SSTTTPVQSGANVVKLGGGGNTLTPVNNLVNASDTCAPTQLLTESPPPLSKTNKK 239
      |||||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 158 -----MHPGGNWL-----GGGQ----- 169
      |||||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 240 ARKKSLPASQPPVAEQVETVLTIETADNIIQAGNLLIVQSPGGQVAVVQVQVPP 299
      |||||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 170 --GGGDLQ-----GTLSTGPAPPLNPOL-----PY-----PS----- 197
      |||||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 300 KAEQQQVQVQIPOQALRVVQAASATLPTVPQKPSQNFQIQAAEPTPTQVYIRTPSGEVTV 359
      |||||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 198 -----DFAPLNPAPY-----APHLLQPGQH---VLPODYKPKAVG-----NS 234
      |||||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 360 LVQDSPPTAATNSNTTCCSPASRAPHLSTGSKKHSAILRKERPLPIAPAGSIISLNA 419
      |||||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 235 GLEGSGAAK-----PPRGAGTGG-----SGG----- 256
      |||||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 420 AQLAAAAQAQMTININGVQGVQVPTITNTGGQQLTVQNVSGNNLTISLSPTQIQLOM 479
      |||||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 257 -----YAGSGAGRSTCDPCNOELERLGAAMAGLRKAPIHSCHIPCGKVYKASHL 308
      |||||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 480 EQALAGETQPGKRRRMACTCPNCKD-----GEKRSGEQGGKKKVVCHIPDCGTFKRTSLL 535
      |||||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 309 KAHLRWHGTGERFVNCNWLFCCKRTRTSDLELHVHVTRETKKFTCLLCKSKRTRSDHLSK 368
      |||||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 536 RAHVLHGTGERFVNCNWFCECKRTRTSDLEQLRHARTHTGDKRFCAQCOKRFRMRSDHLTK 595
      |||||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 369 HORTH 373
      |||||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
596 HYKTH 600
```

Db	107	ASSPSSGRAP	-----GAAPSAAKS	-HRCPPDCAKAYKSHLKLRTHT	152
QY	317	GERPFVCMWLF	CGKRFRTSDE	LERHVHTREKK	TCLLCRKFRFTSDHLSKHQRTHGEP
Db	153	GERPFACDMQ	CGDKKFA	RSDERLARH	RTHTGKRFSCPLCSKFRFTSDHLAKHARRH
QY	377	G	-----PGPPSPGK	ELGEGSVGEE	ANQP-PRSTSPAP
Db	211	GFHPDLRL	RRPGARST	SPSD-----	SLPCSLAGSPAPSPAPSPAP

RESULT 10

ID	KLFD_HUMAN	STANDARD;	PRT;	288 AA.
AC	Q9Y2Y9;	Q9Y356;		
DT	15-JUN-2002	(Rel. 41, Created)		
DT	15-JUN-2002	(Rel. 41, Last sequence update)		
DT	15-JUN-2002	(Rel. 41, Last annotation update)		
DE	Krueppel-like factor 13 (transcription factor BTEB3) (Basic			
DE	transcription element binding protein 3) (BTE-binding protein 3)			
DE	(RANTSS factor of late activated T lymphocytes-1) (RFLAT-1)			
DE	(transcription factor NSLPI) (Novel Spl-like zinc finger transcription			
DE	factor 1) (transcription factor NSLPI).			
GN	KLF13 OR BTEB3 OR NSLPI.			
OS	Homo sapiens (Human).			

RESULT 9

ID	BTE4_HUMAN	STANDARD;	PRT;	252 AA.
AC	Q9BXK1;			
DT	15-JUN-2002	(Rel. 41, Created)		
DT	15-JUN-2002	(Rel. 41, Last sequence update)		
DT	15-JUN-2002	(Rel. 41, Last annotation update)		
DE	transcription factor BTEB4 (Basic transcription element binding-			
DE	protein 4) (BTE-binding protein 4) (Krueppel-like factor 16) (Novel			
DE	Spl-like zinc finger transcription factor 2) (Transcription factor			
DE	NSLP2).			
GN	KLF16 OR BTEB4 OR NSLP2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
ON	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Pancreas;			
RC	Conley A., Urrutia R.;			
RT	"Isolation of a novel zinc finger transcription factor from the			
RT	pancreas extends the repertoire of Spl-like proteins present in this			
RT	organ (Abstract #153)."			
RL	Pancreas 21:437-437(2000).			
CC	-I- FUNCTION: Transcription factor that binds GC and GT boxes and			

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99146379; PubMed=10023774;
 RA Song A., Chen Y.F., Thamatrakoln K., Storm T.A., Krensky A.M.;
 RT "RFLAT-1: a new zinc finger transcription factor that activates RANTES
 gene expression in T lymphocytes.";
 RL Immunity 10:93-103(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99344343; PubMed=10415854;
 RA Cook T., Gebelein B., Urrutia R.;
 RT "Spl and its like: biochemical and functional predictions for a
 growing family of zinc finger transcription factors.";
 RL Ann. N.Y. Acad. Sci. 880:94-102(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21453359; PubMed=11477107;
 RA Kaczynski J., Zhang J.S., Ellenrieder V., Conley A., Duenes T.;
 RA Kester H., van Der Burg B., Urrutia R.;
 RT "The Spl-like protein PTEB3 inhibits transcription via the basic
 transcription element box by interacting with mSin3A and HDAC-1
 co-repressors and competing with Spl.";
 RL J. Biol. Chem. 276:36749-36756(2001).
 CC -!- FUNCTION: Represses transcription by binding to the BTE site, a
 GC-rich DNA element, in competition with the activator Spl. It
 also represses transcription by interacting with the co-repressor
 Sin3A and HDAC1. Activates RANTES expression in T cells.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- TISSUE SPECIFICITY: Ubiquitous.
 CC -!- DOMAIN: The Ala/Pro-rich domain may contain discrete activation
 and repression subdomains and also can mediate protein-protein
 interactions.
 CC -!- PTM: Phosphorylated.
 CC -!- SIMILARITY: BELONGS TO THE SP1 FAMILY OF C2H2-TYPE ZINC-FINGER
 PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AF132599; AAD26864.1;
 DR EMBL: AF150628; AAD34020.1;
 DR EMBL: BC013946; AAH13946.1;
 DR EMBL: BC010438; AAH10438.1;
 DR EMBL: BC012741; AAH12741.1;
 DR HSSP: P08047; 1SP2.
 DR TRANSFAC: T05051;
 DR Genew: HGNC:13672; KLF13.
 DR MIM: 605328;
 DR InterPro: IPR000822; Znf_C2H2.
 DR Pfam: PF00096; zf-C2H2; 6.
 DR ProDom: PD000003; Znf_C2H2; 1.
 DR SMART: SM00355; Znf_C2H2; 3.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 3.
 DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 3.
 KW Transcription regulation; Activator; Repressor; DNA-binding;
 KW Nuclear protein; Repeat; Zinc-finger; Metal-binding; Phosphorylation.
 FT DOMAIN 2 145 ALA/PRO-RICH. (BASIC).
 FT DOMAIN 148 168 ARG/LYS-RICH. (BASIC).
 FT DOMAIN 167 249 ZINC FINGERS.

FT ZN_FING 167 191 C2H2-TYPE.
 FT ZN_FING 197 221 C2H2-TYPE.
 FT ZN_FING 227 249 C2H2-TYPE.
 FT DOMAIN 264 287 SER-RICH.
 FT DOMAIN 2 5 POLY-ALA.
 FT DOMAIN 113 116 POLY-ALA.
 FT CONFLICT 39 39 A -> S (IN REF. 2).
 FT CONFLICT 104 114 EPTSPGAGGAA -> MSPPPPPALKARR (IN REF. 2).
 SQ SEQUENCE 288 AA; 31180 MW; DD2765EE0E9C049 CRC64;
 Query Match 16.6%; Score 395.5; DB 1; Length 288;
 Best Local Similarity 33.6%; Pred. NO. 1.6e-15;
 Matches 114; Conservative 27; Mismatches 95; Indels 103; Gaps 14;
 QY 113 HSSDCLPSVYTSLDWTHPYGWSYKAGIHAGISPGPGNTPTPWMDHMGNNLGGGQGG 172
 DB 9 HFAECLVSMSS-----RAVHV-GREGPESRP-----EG 37
 QY 173 DGLOGLTSTGPAQPLNPLPTYP-----SDFAPL-----NPAPYPAPHLL 213
 DB 38 RAAVAAT-----PTLPRVEERRDCKDSASLEFVARILADLNQAPAPAPAE 83
 QY 214 QGQPOH-----VLPQDVYKAVGNSQLEGSGAA-----KPRGAGTGGGGY 257
 DB 84 REGAAARKARTPCRLPPPAPEPTSPG-----AEGAAAAPPSPAWSEPEAGLEPEREPGP 139
 QY 258 AGSGAGRSTCDPCNCELELGAAGL-----RKKPIHSCHIPGCGKVGKASHLKAHLR 313
 DB 140 AGSGE-----PGLRQVRGRSRADLESPOK---HKCHYAGCEKVGKSHLKAHLR 189
 QY 314 WHTGERPFVCMWFCGKRFTSRDELEHRVHTREKKTCLCSKRFTSDLSLKHQRTH 373
 DB 190 THTGERPFACSWQDCKKRFARSDLELHRYHTGKFKSCPCICEKRFMSDHLTKHARRH 249
 QY 374 G-EGCGPPPGPPELGEGRSGVGEENQPPRSSTSPA 410
 DB 250 ANFHFGMLQRGGSGRTSLSDYSRDSASP---TISPA 285

RESULT 11
 KLF1_HUMAN
 ID KLF1_HUMAN STANDARD; PRT; 416 AA.
 AC Q9UIH9;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE 15-JUN-2002 (Rel. 41, Last annotation update)
 GN Krueppel-like factor 15 (Kidney-enriched krueppel-like factor).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND SUBCELLULAR LOCATION.
 RC TISSUE=Kidney;
 RX MEDLINE=20440192; PubMed=10982849;
 RA Uchida S., Tanaka Y., Ito H., Saitoh-Ohara F., Inazawa J.,
 RA Yokoyama K.K., Sasaki S., Marumo F.;
 RT "Transcriptional regulation of the C/EBP-K promoter by myc-associated
 zinc finger protein and kidney-enriched Krueppel-like factor, a novel
 zinc finger repressor.";
 RL Mol. Cell. Biol. 20:7319-7331(2000).
 CC -!- FUNCTION: Transcriptional activator. Binds to the GA element of
 the C/EBP-K promoter.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- TISSUE SPECIFICITY: Highly expressed in liver, followed by heart,
 skeletal muscle, and kidney. Not expressed in bone marrow or
 lymphoid tissues.
 CC -!- SIMILARITY: BELONGS TO THE SP1 FAMILY OF C2H2-TYPE ZINC-FINGER
 PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -


```
QY 347 REKKFTCLCSKRFTSRDHLKSHQTH 373
      :| | | | | | | | | | | | | | |
Db 327 GHRPFOCHLCDRAFSSDHLALHMKRH 353

RESULT 13
BTE4_MOUSE STANDARD; PRT; 251 AA.
AC P58334;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Transcription factor BTE4 (Basic transcription element binding-
DE protein 4) (BTE-binding protein 4) (krueppel-like factor 16) (Dopamine
GN KLF16 OR BTE4 OR DREF.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RT TISSUE=Neuroblastoma;
MEDLINE=2130923; PubMed=11390978;
Hwang C.K., D'Souza U.M., Eisch A.J., Yajima S., Lammers C.-H.,
Yang Y., Lee S.-H., Kim Y.-M., Nestler E.J., Mouradian M.M.;
RT "Dopamine receptor regulating factor, DRRF: a zinc finger
RT transcription factor";
Proc. Natl. Acad. Sci. U.S.A. 98:7558-7563(2001).
CC -!- FUNCTION: Transcription factor that binds GC and GT boxes in the
CC D1A, D2 and D3 dopamine receptor promoters and displaces Sp1 and
CC Sp3 from these sequences. It modulates dopaminergic transmission
CC in the brain by repressing or activating transcription from
CC several different promoters depending on cellular context.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: High expression in brain; olfactory tubercle,
CC olfactory bulb, nucleus accumbens, striatum, hippocampal CA1
CC region, amygdala, dentate gyrus and frontal cortex. Moderate
CC expression in hippocampal CA2-3 regions, piriform cortex, septum,
CC and distinct thalamic nuclei. Low expression in the cerebellum.
CC -!- DOMAIN: The Ala/Pro-rich domain may contain discrete activation
CC and repression subdomains and also can mediate protein-protein
CC interactions.
CC -!- SIMILARITY: BELONGS TO THE SP1 FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF283891; AAK66968.1;
CC MGD; MGI:2153049; Klf16.
CC TRANSFAC; T05053;
CC InterPro; IPR000822; Znf_C2H2.
CC Pfam; PF00096; Zf-C2H2; 3.
CC SMART; SM00355; Znf_C2H2; 3.
CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
CC PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.
CC Transcription regulation; DNA-binding; Nuclear protein; Repeat;
CC Zinc-finger; Metal-binding.
CC DOMAIN 3 136 ALA/PRO-RICH.
CC FT DOMAIN 103 116 SER-RICH.
CC FT DOMAIN 126 208 ZINC_FINGERS.
CC FT ZN_FING 126 150 C2H2-TYPE.
CC FT ZN_FING 156 180 C2H2-TYPE.
CC FT ZN_FING 186 208 C2H2-TYPE.
CC FT DOMAIN 223 248 PRO/SER-RICH.
CC SEQUENCE 251 AA; 25665 MW; 3F0D7739B1A09FA4 CRC64;
```

```
Query Match 15.9%; Score 379; DB 1; Length 251;
Best Local Similarity 36.8%; Pred No. 1.2e-14;
Matches 105; Conservative 17; Mismatches 93; Indels 70; Gaps 12;

QY 138 AGIHAGISPGPGNTPTPMDMHPGNWLGQGGQDGLQGTSTGPAQPLNPQPTYPY 197
      :| | | | | | | | | | | | | | |
Db 23 AVVHRG-RPGP-----EGAGPAAGLDVRRATREATPPGTPGAPPPPA 63

QY 198 DEAPLNPAYP-----APHLLQPGQHVLVDYKPKAVGNSGQLEGSAAAPPRGAGTG 252
      :| | | | | | | | | | | | | | |
Db 64 -----TAPGGGATAAPHLL-----AASILADLRG-GPVVATAASTAG 100

QY 253 GSGYGAGSAGRSTDCPCNCQELERLGAAGLRKKPIHSCHPGCGKYVGKASHLKAHL 312
      :| | | | | | | | | | | | | | |
Db 101 GTSPVSSSSAASS-----PS-----SGRAPGAAS--HRCFPHGCAKAYKSSHLKSHL 147

QY 313 RWHGTGERPFVPCNWLFCGKRFTSRDELRHVTRHTREKFTCLCSKRFTSRDHLKSHQTH 372
      :| | | | | | | | | | | | | | |
Db 148 RHTHTGERPFACDWPDCDKKFAKRSDELARHRTHTGKFRPCPLCTKRFRSDHLTKHARR 207

QY 373 HCEPGPGPPPS-----GPKELGEGRSVGEEANQP-PRSTSTSPAP 411
      :| | | | | | | | | | | | | | |
Db 208 H-----GFRPELLRRPGARNSVPSDSLXCSLAGSPTSPVPSPAP 248

RESULT 14
KLF2_RAT STANDARD; PRT; 351 AA.
ID KLF2_RAT
AC Q9ET58;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Kruppel-like factor 2 (lung kruppel-like factor).
GN KLF2 OR KLF.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BN/SSNHsd;
RA Haag F., Bartles K., Rothenburg S., Stahmer I., Thiele H.-G.,
RT Koch-Nolte F.;
RT "The gene for the transcription factor KLF is developmentally
RT expressed in rat T cells and is not defective in lymphopenic
RT diabetes-prone BB rats.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: BINDS TO THE CACCC BOX IN THE BETA-GLOBIN GENE PROMOTER
CC AND ACTIVATES TRANSCRIPTION (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF181251; AAG02141.1;
CC HSP; P08047; LSP2.
CC InterPro; IPR000822; Znf_C2H2.
CC Pfam; PF00096; Zf-C2H2; 3.
CC ProDom; PD000003; Znf_C2H2; 2.
CC SMART; SM00355; Znf_C2H2; 3.
CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
CC PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.
CC Transcription regulation; Activator; Zinc-finger; Metal-binding;
CC DNA-binding; Nuclear protein; Repeat.
CC FT DOMAIN 60 70 POLY-PRO.
CC FT DOMAIN 165 168 POLY-PRO.
```

FT DOMAIN 224 227 POLY-ALA.
FT DOMAIN 268 350 ZINC FINGERS.
FT ZN_FING 268 322 C2H2-TYPE.
FT ZN_FING 298 322 C2H2-TYPE.
FT ZN_FING 328 350 C2H2-TYPE.
SQ SEQUENCE 351 AA; 37313 MW; 254EB6B0577A53A4 CRC64;

Query Match 15.7%; Score 374; DB 1; Length 351;
Best Local Similarity 33.8%; Pred. No. 2.9e-14;
Matches 115; Conservative 24; Mismatches 109; Indels 92; Gaps 16;

QY 74 PAGESPPASGVANDYPPPHSFGCP---TGAQDPGLLVKPGHSSDCLPSVYTSMDWTH 130
DB 63 PPQPPPAFYPEGAPPPVGTGPAAGLGTLLRDLDAQO-----PALHGRF-LIA 114

QY 131 PYGSWYKA-----GIHAGISPGPGNTPTPMDMHGPNWGLGGOGDGLQGTLS 180
115 PPGRLVKAEPPEVDGGYGAAGLARG-----PRGLKLEGAL-----GATGACM 158

QY 181 TQPA-QPPLNQLPYPSDFAPLNP-APYPAPHLLOPGQHVLPDQV----- 225
DB 159 RGPAPRPP-----PPSDTTPPLSDGP---PRLPAPGPRNPPFPFPFGPSFGGPGPAL 208

QY 226 -YKRAVGNNGOLESGAA---KPP--RGAGTGGSGVAGSGAGRSTCDPCNCELERLG 279
DB 209 HYGPPAPAGAGFLDDAAALGLAPATRLT-----PPSSPLELE 250

QY 280 A-AAAGLRKKP-----IHSCHIPGCGKVGKASHLKAHLRHTGPRFVCNWLFCGKRFT 333
DB 251 AKPKRGRSRWPKRAATHCTCSYTCNGKTYTKSSHLKAHLRHTGKPYHCNWDGCGWKA 310

QY 334 RSELELHVHTRREKKTCLCSKRFTRSRDHLSKHQTH 373
DB 311 RSEDLTHRYKHTGRHPFQCHLDFRFSRDLALHMRH 350

RESULT 15
KLFF_MOUSE
ID KLFF_MOUSE STANDARD; PRT; 415 AA.
AC Q9EPW2;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Kruppel-like factor 15 (Cardiovascular Kruppel-like factor).
~N KLFI5 OR CKLF.
Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/c;
RA Gray S.J., Kuo C.T., Leiden J.M., Jain M.K.;
RT "CKLF, a cardiovascular Kruppel-like factor."
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Bozell D., Bojunga N., Carninci P., de Bonaudo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Nombarts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE SPI FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).

CC
CC EMBL; AF317225; AAC38597.1; -
CC EMBL; AK009739; BAB26470.1; -
CC EMBL; BC013486; AAH13486.1; -
CC HSSP; P08047; 1SP2.
CC MGD; MGI:1929988; Klf15.
CC InterPro: IPR000822; Znf.C2H2.
CC Pfam; PF00096; zf-C2H2_3.
CC ProDom; PD000003; Znf.C2H2; 1.
CC SMART; SM00355; Znf.C2H2; 3.
CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
CC PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.
CC Transcription regulation: DNA-binding; Activator; Nuclear protein;
CC Repeat; zinc-finger; Metal-binding.
FT ZN_FING 320 344 C2H2-TYPE.
FT ZN_FING 350 374 C2H2-TYPE.
FT ZN_FING 380 402 C2H2-TYPE.
SQ SEQUENCE 415 AA; 44252 MW; 127A7B80DB3E33CC CRC64;

Query Match 15.3%; Score 364; DB 1; Length 415;
Best Local Similarity 25.9%; Pred. No. 1.2e-13;
Matches 123; Conservative 31; Mismatches 120; Indels 200; Gaps 16;

QY 1 MASSLLEEAHYGSSPLAMLT---AACSKF-GGSSP-----LRDSTTLGKGTTK 46
DB 34 MLPSPISDDSDVSSPCSCASPDSCYSCYAGPAGPAGSILDFLLSRATLGGSGSG 93

QY 47 PYADLSAPKTMGDYAPAPFSTNGLLSPAGSPAPASGYANDYPPPHSFGPTGAQDPG 106
DB 94 GIGDSSGPTWGSWRA-----SVPVKEHF-----FPEFLSGDTD----- 130

QY 107 LLVPKGHSSDCLPSVYTSMDWTHPY-----GSWYKAGIHAGISPGPN----- 150
DB 131 -----DVSRFPQPTLEEIEEFLEENMEAEVKEAPENGSRDLETCS 170

QY 151 -----TPTWDMHMGNNLGGGGGOGD----- 174
DB 171 QLSAGSHRSHLHPESAGRERTCP-----PGTSGGGAQAGAPHAHDPVPLVLIQTP 224

QY 175 ---LQGLTSTGPAQPLNPQ-----LP-----TYPSPDFAPLNPAP 206
DB 225 VAVKQAGTGPASPGQAPESVKVQALLVNIQOTFALLPQVVPVSNLNLPSKFRVRIAPV 284

QY 207 YPAPHL-----LQPGPHVLPQDVYKPAVNGSQLEGGAAPKPRGAGTGGGGTAGSA 262
DB 285 IAAKPIGSGSLGPG-----AGLLVGKFKPNP----- 312

QY 263 GRSTCDPCNCELERLGAAGLRKKPHTSCHIPCCKGVYKASHLKAHLRHTGPRPV 322
DB 313 -----AAELK--MHKCTFPCKSMYTKSHLKAHLRHTGPRPV 351


```
QY 61 YPAPESSTNGLSPAGSPAPASGYANDYPPPHSPGPTGAQDPGLLVKGHSSDCLP 120
|||||
Db 61 YPAPESSTNGLSPAGSPAPASGYANDYPPPHSPGPTGAQDPGLLVKGHSSDCLP 120
|||||
QY 121 SVYTSILDTHPYGSKYKAGIHAGISPGGNTPTPMDMHPGNGWLGSGOGDGLQGTLS 180
|||||
Db 121 SVYTSILDTHPYGSKYKAGIHAGISPGGNTPTPMDMHPGNGWLGSGOGDGLQGTLS 180
|||||
QY 181 TGAQPPPLNPOLPTYPSPDFAPLNPAPYPAPHLLOPQPOHVLPOQVYKPKAVGNSQLEGS 240
|||||
Db 181 TGAQPPPLNPOLPTYPSPDFAPLNPAPYPAPHLLOPQPOHVLPOQVYKPKAVGNSQLEGS 240
|||||
QY 241 GAAKPPRGAGTGGSGYAGSAGRSTCDPCNCELERLGAAGLRKPIHSCHIPGCGK 300
|||||
Db 241 GAAKPPRGAGTGGSGYAGSAGRSTCDPCNCELERLGAAGLRKPIHSCHIPGCGK 300
|||||
QY 301 VYKASHLKAHLRWHITGERPFVNCWLFCKRFTSRDELEHVRHTTRKKFTCLCSKRF 360
|||||
Db 301 VYKASHLKAHLRWHITGERPFVNCWLFCKRFTSRDELEHVRHTTRKKFTCLCSKRF 360
|||||
QY 361 TRSDHLSKHQTHGEPGPPSPGPKELGEGRSVGEENANQPPRSTSPAPPEKAHGGSP 420
|||||
Db 361 TRSDHLSKHQTHGEPGPPSPGPKELGEGRSVGEENANQPPRSTSPAPPEKAHGGSP 420
|||||
QY 421 EOSNLEI 428
|||||
Db 421 EOSNLEI 428
|||||

RESULT 2
Q8TDD2 PRELIMINARY; PRT; 431 AA.
AC Q8TDD2
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DE Osterix.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=OSTEOSARCOMA;
RA Ganss B.W.;
RT "cDNA sequence, gene structure and chromosomal localization of the
RT human osterix (OSX) gene."
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF477981; AAL64281.1; --
SQ SEQUENCE 431 AA; 44994 MW; 454A6FEA84309FF9 CRC64;

Query Match 95.0%; Score 2264.5; DB 4; Length 431;
Best Local Similarity 94.7%; Pred. No. 2.6e-157;
Matches 408; Conservative. 4; Mismatches 16; Indels 3; Gaps 1;

QY 1 MASSILLEEEHYGSSPLAMLTAAACSKFGGSPLRDSTTLGKGKTKPY---ADLSAPKTM 57
|||||
Db 1 MASSILLEEEHYGSSPLAMLTAAACSKFGGSPLRDSTTLGKGKTKPY---ADLSAPKTM 57
|||||
QY 58 GDAYPAPTSSTNGLSPAGSPAPASGYANDYPPPHSPGPTGAQDPGLLVKGHSSD 117
|||||
Db 61 GDAYPAPTSSTNGLSPAGSPAPASGYANDYPPPHSPGPTGAQDPGLLVKGHSSD 120
|||||
QY 118 CLPVSVYTSILDTHPYGSKYKAGIHAGISPGGNTPTPMDMHPGNGWLGSGOGDGLQGT 177
|||||
Db 121 CLPVSVYTSILDTHPYGSKYKAGIHAGISPGGNTPTPMDMHPGNGWLGSGOGDGLQGT 180
|||||
QY 178 TLSTGAPAPPLNPOLPTYPSPDFAPLNPAPYPAPHLLOPQPOHVLPOQVYKPKAVGNSQ 237
|||||
Db 181 TLPTGAPAPPLNPOLPTYPSPDFAPLNPAPYPAPHLLOPQPOHVLPOQVYKPKAVGNSQ 240
|||||
QY 238 EGSGAAGPPRGAGTGGSGYAGSAGRSTCDPCNCELERLGAAGLRKPIHSCHIPG 297
|||||
```

```
Db 241 EGSGAAGPPRGAGTGGSGYAGSAGRSTCDPCNCELERLGAAGLRKPIHSCHIPG 300
|||||
QY 298 CGKYVGKASHLKAHLRWHITGERPFVNCWLFCKRFTSRDELEHVRHTTRKKFTCLCS 357
|||||
Db 301 CGKYVGKASHLKAHLRWHITGERPFVNCWLFCKRFTSRDELEHVRHTTRKKFTCLCS 360
|||||
QY 358 KRFTSRDHLKHQTHGEPGPPSPGPKELGEGRSVGEENANQPPRSTSPAPPEKAH 417
|||||
Db 361 KRFTSRDHLKHQTHGEPGPPSPGPKELGEGRSVGEENANQPPRSTSPAPPEKAH 420
|||||
QY 418 GSPEQSNNLEI 428
|||||
Db 421 GSPEQSNNLEI 431
|||||

RESULT 3
Q96MJ1 PRELIMINARY; PRT; 452 AA.
AC Q96MJ1
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE CDNA FLJ32295 fis, clone PROST2001823, weakly similar to transcription
DE factor Sp1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PROSTATE;
RA Ishibashi T.; Kanehori K.; Yosida M.; Watanabe S.; Ishida S.; Ono Y.;
RA Hotuta T.; Hiraoka S.; Murakawa K.; Takiguchi S.; Kusano J.;
RA Watanabe M.; Fujimori K.; Tanai H.; Ishida M.; Yamashita H.; Chiba Y.;
RA Sugiyama T.; Irie R.; Otsuki T.; Sato H.; Wakamatsu A.; Ishii S.;
RA Yamamoto J.; Isono Y.; Kawal-Hio Y.; Saito K.; Nishikawa T.;
RA Kimura K.; Matsuo K.; Nakamura Y.; Sekine M.; Kikuchi H.; Kanda K.;
RA Wagatsuma M.; Takahashi-Fujii A.; Oshima A.; Sugiyama A.; Kawakami B.;
RA Suzuki Y.; Sugano S.; Nagahari K.; Masuho Y.; Nagai K.; Isogai T.;
RT "NEDO human cDNA sequencing project."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK056857; BAB71297.1; --
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; Zf-C2H2; 3.
DR ProDom: PD000003; Znf_C2H2; 2.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_3.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 3.
KW DNA-binding; Zinc-finger.
SQ SEQUENCE 452 AA; 45836 MW; 95383C60C112320F CRC64;

Query Match 34.3%; Score 818.5; DB 4; Length 452;
Best Local Similarity 42.4%; Pred. No. 6.7e-52;
Matches 196; Conservative. 50; Mismatches 155; Indels 61; Gaps 17;

QY 1 MASSILLEEEHYGSSPLAMLTAAACSKFGGSP---LRD-STTLGKG-----GKPKPYA 49
|||||
Db 1 MATSLLEEEHYGSSPLAMLTAAACSKFGGSPSSLSGSGFGKFPWKRSSSSSA 60
|||||
QY 50 DLSAPKTMGDYAPFSTNGLSPAGSPAPASGYANDYPPPHSPGPTGAQDPG--- 106
|||||
Db 61 SCNVVGSLSGFGVSGASRNGSSSAAAAAALVSDSFSCGSSAHSQDSGSHQ 120
|||||
QY 107 -LLVPKGHSSDCLPVSVYTSILDTHPYGSKYKAGIHAGI-----SPGNGTPTPMDM 158
|||||
Db 121 PVFTSKVHTSDVGLQGTIPRYGMARHPVSWFKPS-HPGGLGAGEVGSAGASS---WWDV 175
|||||
QY 159 HPGGNWLG-GGQGGDGLQGT---STGPAQPLNPOLPTYPSPDFAPLNPAPY---PAPHL 212
|||||
Db 176 --GAGWDVQNPNSAAALPGSLHPAAGGLQTLHSPGLGYNDSYGLSHSFASSGASHL 233
|||||
QY 213 LQPGFQHVLPQDVYKPKAVGNSQLEGSAAKPPRGAGTGG-----GGVAGSGA 262
|||||
```


Db 237 LPATIDGT-GVLTQATVAVSGVSD-----PPSTENFNHNLQO--MOVSSSATTTISQP 288
 QY 121 SVYTSLDMTHPYGSWKAGIHAGIS-PCPGNTPTPMDHMPGGNWL-----GGGQGG 171
 Db 289 ILQLSGD-----GQAQVAGQDLGAGTMSQVLVNPFGTLIAQATVTTATGQIQWQ 340
 QY 172 GDGLQGLTSLGPAQPLN-----POLPYPSDFAPLNA-PYPAPHL-----LQPGPQ 218
 Db 341 TFQVGQVOSLQGLQPOGQQAQOQLTAPVQTLPLGAGQVSLPNLQATVTVNSVTQTGVQ 400
 QY 219 HVLQDVYKPAV-----GNSQLEGSGAAKP-----PRGACTGGGGVAGSAG 263
 Db 401 YAQGEARSPSGIOIKKEPDPSEEWQLSGDSTLNPDSLNLRGPMGDEDMEAAGDGKRLR 460
 QY 264 RSTDCPCNCOELERLGAAGLRKKPIHSCHIPGCGKVYKASHLKAHLRMTGERPFVC 323
 Db 461 RVACTCPNCKE---SGRGSGVGKKQHICHIPGCGKVYKASHLKAHLRMTGERPFVC 517
 QY 324 NWLFCGRKFTSRDELEHVRTHTRKKFTCLLCKSRKTRSDHLSKHORTH-GEPCGCPPPP 382
 Db 518 NMTYCGKRFTRSDQLQRHRRTHTEKRFVCSKCRFMRSDHLAKTIKTHQNKGGGPPS 577
 QY 383 SGP 385
 Db 578 TSP 580
 RESULT 8
 ID Q90XF6 PRELIMINARY; PRT; 367 AA.
 AC Q90XF6;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Zinc finger buttonhead-related transcription factor 1.
 GN Bts1.
 OS Brachydanio rerio (Zebrafish) (Zebra danio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 RN NCBI_TaxID=7955;
 RX SEQUENCE FROM N.A.
 RP MEDLINE=1521131; PubMed=116411235;
 RA Tallatius A., Wilms T.P., Crozatier M., Pfeiffer P., Wassef M.,
 Bally-Cuif L.;
 RT "The zebrafish buttonhead-like factor Bts1 is an early regulator of
 pax2.1 expression during mid-hindbrain development.";
 RL Development 128:4021-4034(2001).
 DR EMBL; AF388363; AAK83353.1;
 DR InterPro; IPR000822; Znf_C2H2.
 DR Pfam; PF00096; zf-C2H2; 3.
 DR ProDom; PD000003; Znf_C2H2; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_3.
 DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 3.
 DR DNA-binding; Zinc-finger.
 SQ SEQUENCE 367 AA; 41030 MW; F31946B6A66CD8FD CRC64;
 Query Match 20.1%; Score 480; DB 13; Length 367;
 Best Local Similarity 31.7%; Pred. No. 2.5e-27;
 Matches 133; Conservative 34; Mismatches 100; Indels 152; Gaps 20;
 QY 7 EEEAHYSSPLAMITAAACSKFG-----GSSP-----LRDSTTLGKG----- 43
 Db 26 ENSKH---SPLALLAATCNRIHGHGHTPTDFIQVYDITLGSRSRIFHPWNSNEANHQST 82
 QY 44 -TKPYADLSAPKWTGDAYPA-----PFSSTNGLLS-----PAGSP--A 80
 Db 83 LSSNPSFGLSKSLQSSYASHHELPLTPADPTYPIMISLRCHVNASLQSTCPTTV 142
 QY 81 PASGVANDYPPFPHSGPTGAQDGLLVKPGHSSDCLPSVYTSLDMTHPYGSWKAGI 140
 Db 143 PAVTYAP-APIPPAMS-----FVP-GHSG-----LVHQO----- 171

QY 141 HAGISFPGNTPTPMDHMPGGNWLGGGQGGDLQGLTSLGPAQPLNQLPTYPSPDEFA 200
 Db 172 QRQLSPNPGD-DIPWNSLQGN----- 192
 QY 201 PLNAPYPAHLLQPGQHVLPQDVYKPAVGN--COLESGAA-----KPPRGAGTGGG 254
 Db 193 -----PVAHSVHP-----HRFP--IQRLVLGHTDFAYQYQIAALLHTKSP----- 232
 QY 255 GGYAGSAGRSTCDPCNCOELERLGAAGLRKKPIHSCHIPGCGKVYKASHLKAHLRW 314
 Db 233 ---LATARRCRRRCPCNQS-----SSSDEFGKKQHICHIPGCGKVYKASHLKAHLRW 285
 QY 315 HTGERPFVCMWLCGRKFTSRDELEHVRTHTRKKFTCLLCKSRKTRSDHLSKHORTH 373
 Db 286 HSGERPFVCMWLCGRKFTSRDELEHVRTHTRKKFTCLLCKSRKTRSDHLSKHORTH 344
 RESULT 9
 ID Q8UUU3 PRELIMINARY; PRT; 373 AA.
 AC Q8UUU3;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Spl-1-like zinc-finger protein XSPR-1.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 OC Xenopodinae; Xenopus.
 RN NCBI_TaxID=8355;
 RX SEQUENCE FROM N.A.
 RP Ossipova O., Stick R., Pieler T.;
 RT "Interaction between two novel Spl-1-like zinc finger proteins and
 brachyury in xenopus.";
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY062264; AAL47217.1;
 DR InterPro; IPR000822; Znf_C2H2.
 DR Pfam; PF00096; zf-C2H2; 3.
 DR PRINTS; PR00048; ZINCFINGER.
 DR ProDom; PD000003; Znf_C2H2; 1.
 DR SMART; SM00355; Znf_C2H2; 3.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_3.
 DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 3.
 KW Zinc.
 SQ SEQUENCE 373 AA; 41496 MW; EBFT4F648153CDA CRC64;
 Query Match 19.6%; Score 468; DB 13; Length 373;
 Best Local Similarity 32.8%; Pred. No. 1.9e-26;
 Matches 133; Conservative 36; Mismatches 107; Indels 130; Gaps 23;
 QY 15 SPLAMITAAACSKFG--GSSP-----LRDSTTLGK-----GGTKKPYAD 50
 Db 31 TPLALLAATCNRIHGHGHTPTDFIQVYDITLGSRSRIFHPWNEIPTHTSGGI-OPHNG 89
 QY 51 LSAPKT-MGDYAPAPFSTNGLLSPAGSPAPAPAGYANDYPPPHSPGPTGAQDPLLV 109
 Db 90 IGLQKTLASHIOSSFVTHPLTPPADP-----SYPEYKSPV-KLLPPPMMA----- 136
 QY 110 PKGHSSDCLPSVYTSLDMTHPYG--SWYKAGIHAGISPGPN-----TPT----- 153
 Db 137 ---FOASAC-QSAYVP---TVPYAPPAPITSAH-GFVPSHSLNHQHPMQLSPNLAEDM 188
 QY 154 PWDMDHMPGGNWLGGGQGGDLQGLTSLGPAQPLNQLPTYPSPDFAPLNPAPYAPHLL 213
 Db 189 PWSIQ-----QAT-----PVAHAS-PIPH-- 207
 QY 214 QPGQHVLPQDVYKPAVGN--GSSP-----OLESGAAKPPRGAGTGGSGYAGSAGRSTC 267
 Db 208 ---PHHFRP---IQRLVLGHTDFAYQYQIAALLHTKSP-----LATARRCRR 251
 QY 268 DCPNCOELERLGAAGLRKKPIHSCHIPGCGKVYKASHLKAHLRMTGERPFVCMWLF 327

```

Db 252 RCPNQTSS--GPDEVG--KKQHCHIPGCGKVGKTSHLKAHLRWHSGERPFICWNLF 307
|||||
Qy 328 CGKFRTRSDLELRHVRTHTRKFTCLLCKRFRTRSDHLSKHQTH 373
|||||
Db 308 CGKFRTRSDLELRHVRTHTRKFTCLLCKRFRTRSDHLSKHQTH 353
|||||

RESULT 10
Q64167 PRELIMINARY: PRT; 467 AA.
AC Q64167; Q62251;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Transcription factor 1 (transcription factor Sp1) (SP1
DE gene) (3' end).
GN SPl.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
MEDLINE=96016118; PubMed=7568082;
RA Persengiev S.P., Saffer J.D., Kilpatrick D.L.;
RT "An alternatively spliced form of the transcription factor Sp1
RT containing only a single glutamine-rich transactivation domain.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:9107-9111(1995).
RN [2]
SEQUENCE OF 367-467 FROM N.A.
MEDLINE=92338398; PubMed=1633330;
RA Chestier A., Charnay P.;
RT "Difference in the genomic organizations of the related transcription
RT factors Sp1 and Krox-20; possible evolutionary significance.";
RL DNA Seq. 2:325-327(1992).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL; S79832; BAB35321.1;
DR EMBL; X60136; CAA42721.1;
DR HSSP; P08047; ISPl.
DR InterPro: IPR000822; Znf_C2H2; 3.
DR MGD; MGI:98372; SPl.
DR Pfam; PF00096; zf-C2H2; 3.
DR PRINTS; PR00048; ZINC_FINGER.
DR ProDom; PD000003; Znf_C2H2; 2.
DR SMART; SM00355; Znf_C2H2; 3.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 3.
KW DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
SQ SEQUENCE 467 AA; 48749 MW; 4A373D67E6128197 CRC64;

Query Match 19.5%; Score 464.5; DB 11; Length 467;
Best Local Similarity 43.5%; Pred. No. 4.4e-26;
Matches 111; Conservative 16; Mismatches 77; Indels 51; Gaps 9;

Qy 171 QGDGLOGLTSTGPAQPLNPQLPTYPSPDFAPLNPAPYPAPHLQPGQHVLPQDVYKPKA 230
|||||
Db 173 QGVSLGTSSNTLTPI-ASAASIPAGTVTVNAALSS-----MPGLQTI-----NLSA 221
|||||

Qy 231 VGNSSG----OLEG--SGAARPPRGAGT-----GGSGG-----YAGS 260
|||||
Db 222 LGTSGIQVHOLPGLPLAINTPGDHGTLGLHSGGGGDIHDETAGGEGSSDLQPOAGR 281
|||||

Qy 261 GAGRSTDCPCNCELERLGAAAGLRKKPIHSCHIPGCGKVGKASHLKAHLRWHHTGERP 320
|||||
Db 282 RTRREACTCPYCKDSE--GRASGDPGKKQHICHIOCGKVGKTSHLKAHLRWHHTGERP 339
|||||

Qy 321 FVCNWLFCGRFRTRSDLELRHVRTHTRKFTCLLCKRFRTRSDHLSKHQTHGEPGPGP 380
|||||
Db 340 FMCNWSYCGKRFRTRSDLELRHVRTHTRKFTCLLCKRFRTRSDHLSKHQTHGEPGPGP 399
|||||

Qy 381 PPSGPKELGEGRSVG 395
|||||

```

```

Db 400 -----GVALSVG 406

RESULT 11
O89090 PRELIMINARY: PRT; 781 AA.
AC O89090;
DT 01-NOV-1998 (TEMBLrel. 08, Created)
DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Transcription factor Sp1.
GN SPl.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
TISSUE=NEUROBLASTOMA;
RX MEDLINE=98290594; PubMed=9628590;
RA Yajima S., Lee S.H., Minowa T., Mouradian M.M.;
RT "Sp family transcription factors regulate expression of rat D2
RT dopamine receptor gene.";
RL DNA Cell Biol. 17:471-479(1998).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL; AF062566; AAC16484.1;
DR HSSP; P08047; ISPl.
DR MGD; MGI:98372; SPl.
DR InterPro: IPR000822; Znf_C2H2; 3.
DR Pfam; PF00096; zf-C2H2; 3.
DR PRINTS; PR00048; ZINC_FINGER.
DR ProDom; PD000003; Znf_C2H2; 2.
DR SMART; SM00355; Znf_C2H2; 3.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 3.
KW DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
SQ SEQUENCE 781 AA; 80486 MW; 14CD12BEC58CF921 CRC64;

Query Match 19.5%; Score 464.5; DB 11; Length 781;
Best Local Similarity 43.5%; Pred. No. 7.8e-26;
Matches 111; Conservative 16; Mismatches 77; Indels 51; Gaps 9;

Qy 171 QGDGLOGLTSTGPAQPLNPQLPTYPSPDFAPLNPAPYPAPHLQPGQHVLPQDVYKPKA 230
|||||
Db 487 QGVSLGTSSNTLTPI-ASAASIPAGTVTVNAALSS-----MPGLQTI-----NLSA 535
|||||

Qy 231 VGNSSG----OLEG--SGAARPPRGAGT-----GGSGG-----YAGS 260
|||||
Db 536 LGTSGIQVHOLPGLPLAINTPGDHGTLGLHSGGGGDIHDETAGGEGSSDLQPOAGR 595
|||||

Qy 261 GAGRSTDCPCNCELERLGAAAGLRKKPIHSCHIPGCGKVGKASHLKAHLRWHHTGERP 320
|||||
Db 596 RTRREACTCPYCKDSE--GRASGDPGKKQHICHIOCGKVGKTSHLKAHLRWHHTGERP 653
|||||

Qy 321 FVCNWLFCGRFRTRSDLELRHVRTHTRKFTCLLCKRFRTRSDHLSKHQTHGEPGPGP 380
|||||
Db 654 FMCNWSYCGKRFRTRSDLELRHVRTHTRKFTCLLCKRFRTRSDHLSKHQTHGEPGPGP 713
|||||

Qy 381 PPSGPKELGEGRSVG 395
|||||
Db 714 -----GVALSVG 720

RESULT 12
O89087 PRELIMINARY: PRT; 784 AA.
AC O89087;
DT 01-NOV-1998 (TEMBLrel. 08, Created)
DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Transcription factor Sp1.
GN SPl.
OS Mus musculus (Mouse).

```


